

NASA TECHNICAL  
MEMORANDUM

NASA TM X-53292

July 12, 1965

NASA TM X-53292

FACILITY FORM 802	N65 33050	N65 33064
	(ACCESSION NUMBER)	(THRU)
	488	1
	(PAGES)	(CODE)
	TMX 53292	30
	(NASA CR OR TMX OR AD NUMBER)	(CATEGORY)

PROGRESS REPORT NO. 7  
ON STUDIES IN THE FIELDS OF SPACE FLIGHT AND GUIDANCE  
THEORY

Sponsored by AERO-ASTRODYNAMICS LABORATORY

NASA

*George C. Marshall  
Space Flight Center,  
Huntsville, Alabama*

GPO PRICE \$ \_\_\_\_\_

CSFTI PRICE(S) \$ \_\_\_\_\_

Hard copy (HC) \$1.88

Microfiche (MF) 2.50

ff 653 July 65

TECHNICAL MEMORANDUM X-53292

PROGRESS REPORT NO. 7  
on Studies in the Fields of  
SPACE FLIGHT AND GUIDANCE THEORY

Sponsored by Aero-Astroynamics Laboratory of  
Marshall Space Flight Center

ABSTRACT

The progress reports of NASA-sponsored studies in the areas of space flight and guidance theory are presented. The studies are carried on by several universities and industrial companies. This progress report covers the period from July 23, 1964 to April 1, 1965. The contracts are technically supervised by personnel of the Astroynamics and Guidance Theory Division, Aero-Astroynamics Laboratory, Marshall Space Flight Center.

NASA-GEORGE C. MARSHALL SPACE FLIGHT CENTER

NASA - GEORGE C. MARSHALL SPACE FLIGHT CENTER

---

TECHNICAL MEMORANDUM X-53292

---

PROGRESS REPORT NO. 7  
on Studies in the Fields of  
SPACE FLIGHT AND GUIDANCE THEORY

Sponsored by Aero-Astroynamics Laboratory of  
Marshall Space Flight Center

ASTRODYNAMICS AND GUIDANCE THEORY DIVISION  
AERO-ASTRODYNAMICS LABORATORY  
RESEARCH AND DEVELOPMENT OPERATIONS

# TECHNICAL MEMORANDUM X-53292

## PROGRESS REPORT NO. 7 on Studies in the Fields of SPACE FLIGHT AND GUIDANCE THEORY

### SUMMARY

The progress reports of NASA-sponsored studies in the areas of space flight and guidance theory are presented. The studies are carried on by several universities and industrial companies. This progress report covers the period from July 23, 1964 to April 1, 1965. The contracts are technically supervised by personnel of the Astrodynamics and Guidance Theory Division, Aero-Astrodynamics Laboratory, Marshall Space Flight Center.

### INTRODUCTION

This report contains fourteen papers, the subject matter of which lies in the areas of space flight and guidance theory. These papers were written by investigators employed at agencies under contract to MSFC.

This report is the seventh of the "Progress Reports" and covers the period from July 23, 1964 to April 1, 1965. Information given in the earlier progress reports will not be repeated here.

The agencies contributing and their fields of major interest are:



Field of Interest	Agency
Optimization Theory (Calculus of Variations)	Vanderbilt University Auburn University Analytical Mechanics Associates
Orbital Transfer	North American Aviation, Inc. United Aircraft Corporation
Control Theory	Honeywell, Inc.
Celestial Mechanics	University of Wisconsin Hayes International Corporation
Low Thrust Trajectories	Grumman Aircraft Engineering Corp.
Large Computer Exploitation	Georgia Institute of Technology Southern Illinois University

The objective of this introduction is to briefly review the contributions of each agency.

The first paper by Dr. M. Boyce and Mr. J. Linnstaedter of Vanderbilt University develops a multiplier rule and analogues of the Weierstrass and Clebsch conditions for a multi-stage Bolza-Mayer calculus of variations problem. The number of stages is fixed, but partition points defining stage boundaries are variable. Discontinuities are allowed in variables and constraint functions at partition points. The constraints include finite equations and inequalities, as well as differential equations, all of which involve control variables. An appendix to the report summarizes some of the results obtained by C. H. Denbow, as modified by R. W. Hunt, for a generalized Bolza problem.

The second paper by Joe W. Reece and Grady R. Harmon is an application of the necessary conditions resulting from the Pontryagin Maximum Principle to a particular model for the simulation of reentry trajectories. The paper is a good example of the detailed analysis needed to achieve a workable computational procedure, but the method used to solve for the boundary conditions is yet to be incorporated into the procedure.

The third paper by Henry J. Kelley and Walter F. Denham derives the necessary conditions for optimal guidance polynomial approximations by an ensemble averaging approach. The merits of this approach can better be evaluated when a computational scheme utilizing the derived necessary conditions is outlined and applied to a trajectory analysis problem.

The fourth paper by Gary A. McCue and David F. Bender of North American Aviation presents a method for the numerical determination of optimum two-impulse orbital transfers between inclined elliptical orbits. A numerical optimization technique termed "adaptive steepest descent" is shown to overcome convergence difficulties. Results are obtained for "almost target" coplanar elliptical orbits. Extensions are then developed for strongly inclined orbits.

The fifth paper by David F. Bender and Gary A. McCue of North American Aviation presents numerical and analytical results concerning optimum one-impulse orbital transfer maneuvers. Approximate expressions for the minima of the one-impulse maneuvers are derived. Numerical comparisons of one-impulse transfers and corresponding optimum two-impulse transfers are made. These comparisons show that for a small range of shapes, one-impulse transfers are optimal.

The sixth paper by Frank Gobetz of United Aircraft Corporation presents a study of the minimum fuel transfer and rendezvous between neighboring low-eccentricity orbits by power-limited rockets. The equations of motion are linearized in three separate coordinate systems, given a variational treatment, and solved in closed form. Both performance type and guidance type of solutions are presented in each of the three systems. By choosing an intermediate orbit for the reference orbit in an application of the linear theory to interplanetary transfer, results for Earth-Venus and Earth-Mars transfers are found to agree well with exact results.

The seventh paper, submitted by E. B. Lee of Honeywell, is entitled, "An Approximation to Linear Bounded Phase Coordinate Control Problems." The technique employs a non-negative "penalty function" which is small for state variables satisfying the given constraints, and large outside of this constraint set. An optimal control problem is solved where, as a terminal condition, the integral of the penalty function is bounded by a small constraint, thereby limiting the excursions of the state variables outside of the constraint set.

The eighth paper by C. C. Conley of the University of Wisconsin studies the solutions of the restricted three-body problem near those equilibrium points which are collinear with the two positive masses. This is done to gain insight toward the development of an analytic proof and classification of the periodic orbits that pass near these equilibrium points, which have been discovered numerically by M. Davidson, and also to hopefully gain insight into the nature of solutions of the restricted three-body problem in general. The qualitative observations that are made are all deduced from the linearized equations.

The ninth paper, by A. A. Nafsoosi and H. Passmore of Hayes International Corporation, considers an approach to the analytical solution of the minimum fuel trajectory integration problem through the Hamilton-Jacobi theory of canonical transformations. This method replaces the ordinary differential equations of motion with the Hamilton-Jacobi partial differential equations. The method of separation of variables and Jacobi's method for solving partial differential equations are discussed and applied to progressively more realistic approximations to the minimum fuel trajectory problem. This approach is found to be of limited usefulness unless a more appropriate transformation of the coordinates can be found that would produce a more easily solvable Hamilton-Jacobi equation.

The tenth paper, by Harry Passmore, also applies methods of celestial mechanics to the problem of deriving an analytical solution to the minimum fuel trajectory problem. By considering the  $\lambda$  variables as coordinates of a fictitious body relative to the vehicle, and transforming the  $\lambda$  equations to equations relative to the same center of attraction as the vehicle, equations analogous to the three-body equations are obtained. These equations are transformed to canonical equations and solved by Delaunay procedures. The solution obtained is a first order approximation expected to be most applicable to the many-orbit low-thrust problem rather than interplanetary transfer or high-thrust trajectory integration.

The eleventh paper, by Hans K. Hinz, Robert McGill, and Gerald Taylor, and the twelfth paper by Paul Kenneth and Gerald E. Taylor, relate to their numerical experience with the generalized Newton-Raphson method reported on in Progress Report No. 5, as applied to the low thrust two-point boundary value problem. Equations of motion in both applications are formulated in two-dimensional polar coordinates. One application concerns geocentric circular orbital transfer. Simple

equations are given for first values used to begin the iteration. Successful results have been achieved for trajectories of up to twenty-one revolutions and correct to four significant figures with convergence deteriorating past this point. Greater accuracy may be expected by using multiple precision arithmetic and better numerical integration methods.

The second application of the numerical method concerns the interplanetary trajectory with bounded thrust magnitude and thrust angle used as control variables. Transit time is specified and mass is maximized. Again, convergence has been obtained to an accuracy of four significant figures with further possibilities for improvement by using better numerical methods.

It seems the Generalized Newton-Raphson Method shows value for meeting specified end-conditions for the sensitive low thrust trajectory optimization problem, although the geocentric spiral trajectory sensitivity may still offer difficulty.

The thirteenth paper by I. E. Perlin, J. H. Mackay, et al., contains a very thorough examination of the many different aspects of multivariable function approximation by least squares techniques. It also contains some illuminating examples of the mathematical techniques which are used for the selection of a few efficient estimation variables from a larger set.

The fourteenth paper by Robert Silber describes a procedure for numerically computing the coefficients for the Taylor's series expansion of the general solution of a normal system of first order, ordinary differential equations in terms of the time on any solution and the initial values of the variables and time for that solution. The method has appeared previously in NASA TM X-53059 as part of a more involved method to compute a guidance type of solution for a system of differential equations. The present paper singles out the first mentioned solution as possibly deserving explicit mention, and brings out the mathematical considerations that justify this procedure and that are necessary for one to make reasonable applications of it.

VANDERBILT UNIVERSITY

N65 33051

NECESSARY CONDITIONS FOR A MULTISTAGE  
BOLZA-MAYER PROBLEM INVOLVING CONTROL VARIABLES AND  
HAVING INEQUALITY AND FINITE EQUATION CONSTRAINTS

by

M. G. Boyce

J. L. Linnstaedter

NASHVILLE, TENNESSEE

DEPARTMENT OF MATHEMATICS  
VANDERBILT UNIVERSITY  
NASHVILLE, TENNESSEE

NECESSARY CONDITIONS FOR A MULTISTAGE  
BOLZA-MAYER PROBLEM INVOLVING CONTROL VARIABLES AND  
HAVING INEQUALITY AND FINITE EQUATION CONSTRAINTS

By M. G. Boyce and J. L. Linnstaedter

SUMMARY

33051  
A multiplier rule and analogues of the Weierstrass and Clebsch conditions are developed for a multistage Bolza-Mayer calculus of variations problem. The number of stages is fixed, but partition points defining stage boundaries are variable. Discontinuities are allowed in variables and constraint functions at partition points. The constraints include finite equations and inequalities, as well as differential equations, all of which involve control variables.

An Appendix summarizes some of the results obtained by C. H. Denbow, as modified by R. W. Hunt, for a generalized Bolza problem. The Appendix is independent of the rest of the paper.

## NOTATION

### Ranges of Subscripts and Superscripts

$a = 1, \dots, p$	$e, j = 1, \dots, m$	$\alpha, \eta = 1, \dots, N = n+m+r$
$b = 1, \dots, p-1$	$g = 1, \dots, q$	$\beta, \delta = 1, \dots, M = n+q+r$
$c = 0, \dots, s$	$i = 1, \dots, n$	$\gamma = 1, \dots, K = s+m+r$
$d, h = 1, \dots, r$	$k = 1, \dots, s$	$\rho = 0, \dots, K = s+m+r$

### Intervals, Regions and Arcs

$I$	interval $t_0 \leq t \leq t_p$ .
$I_a$	subinterval between partition points $t_{a-1}$ and $t_a$ .
$R_a$	open connected set in $(t, x, y)$ space.
$S$	open connected set in $(t_0, \dots, t_p, x(t_0), x(t_1^-), x(t_1^+), \dots, x(t_p))$ space.
$R'_a$	open connected set in $(t, z, \dot{z})$ space.
$S'$	open connected set in $(t_0, \dots, t_p, z(t_0), z(t_1^-), z(t_1^+), \dots, z(t_p))$ space.
$C_a$	admissible sub-arc.
$E$	admissible arc.
$C'_a$	admissible sub-arc for transformed and Appendix problems.
$E'$	admissible arc for transformed and Appendix problems.

### Functions and Variables

$t$	independent variable.
$t_0, \dots, t_p$	partition set with $t_0 < t_1 < \dots < t_p$ .
$x$	state variable vector $(x_1, \dots, x_n)$ .
$y$	control variable vector $(y_1, \dots, y_m)$ .
$L^a_i$	differential equation functions, $t$ in $I_a$ .
$M^a_g$	finite equation functions, $t$ in $I_a$ .

$N_h^a$	inequality constraint functions, $t$ in $I_a$ .
$L_i$	differential equation functions, $t$ in $I$ ; $L_i = L_i^a$ , $t$ in $I_a$ .
$M_g$	finite equation functions, $t$ in $I$ ; $M_g = M_g^a$ , $t$ in $I_a$ .
$N_h$	inequality constraint functions, $t$ in $I$ ; $N_h = N_h^a$ , $t$ in $I_a$ .
$J_k$	end and intermediate point constraint functions.
$J_o$	function to be minimized.
$\lambda$	multiplier vector $(\lambda_1, \dots, \lambda_n)$ for differential equations.
$\mu$	multiplier vector $(\mu_1, \dots, \mu_q)$ for finite equations.
$\nu$	multiplier vector $(\nu_1, \dots, \nu_r)$ for inequality constraints.
$D_1^a, D_2^a, A, B$	diagonal matrices.
$x(t_b^-), z(t_b^-)$	left hand limit at $t_b$ .
$x(t_b^+), z(t_b^+)$	right hand limit at $t_b$ .
$d_{ib}, d_{ab}$	amount of discontinuity at $t_b$ .
$c_i^a, c_\alpha^a$	integration constants in multiplier rules.
$z$	vector $(z_1, \dots, z_N)$ for transformed and Appendix problems.
$\phi_\beta^a$	differential equation functions for $z$ -system problems.
$f_\gamma$	end and intermediate conditions for $z$ -system.
$f_o$	function to be minimized for $z$ -system.
$\lambda_\beta$	multipliers for $z$ -system.
$F$	Lagrangian function.
$H$	generalized Hamiltonian function.
$E$	Weierstrass E-function.
$e_\rho$	constant multipliers in transversality equations.
$\pi, \theta$	Clebsch condition variables.
$Z, Y$	Weierstrass condition variables.



## INTRODUCTION

In 1937 C. H. Denbow (reference 1) formulated a multistage generalization of the Bolza problem and established necessary and sufficient conditions for it. His method involved transforming the multistage problem into a standard problem of Bolza by a transformation due to W. T. Reid and L. M. Graves. The transformation requires that the number of stages be fixed and the staging points be distinct. By stages we refer to the subintervals into which the range of the independent variable is partitioned by intermediate points involved in the constraints.

R. W. Hunt (2) has applied Denbow's methods to a Mayer form of the multistage problem in which discontinuities are permitted in the variables and constraints at staging points. He obtained the first three necessary conditions. We have summarized his results in the Appendix, with some minor modifications.

In this paper we further extend the work of Denbow and Hunt to include control variables, finite equation conditions, and inequality constraints. Following Hunt, we use the Mayer formulation, which Bliss (3, p. 190) has shown equivalent to the Bolza form for one stage problems. Also we have used differential constraints in normal form, a form directly applicable to trajectory optimization. Hestenes (4, pp. 4-6) has shown that the one stage problem in this form with control variables is reducible to the usual form of the Bolza problem and vice-versa. The method of Valentine (5) is used to transform the inequality constraints into differential equations.

## FORMULATION OF THE PROBLEM

Let  $t$  be the independent variable. Define a set of variables  $(t_0, \dots, t_p)$  contained in the range of  $t$  to be a partition set if and only if  $t_0 < t_1 < \dots < t_p$ . Call the elements of the partition set partition points. Let  $I$  denote the interval  $t_0 \leq t \leq t_p$ , and let  $I_a$  denote the sub-interval  $t_{a-1} \leq t < t_a$  for  $a = 1, \dots, p-1$  and  $t_{a-1} \leq t \leq t_a$  for  $a = p$ .

Let  $x(t)$  denote the set of functions  $(x_1(t), \dots, x_n(t))$ . For each  $i, i = 1, \dots, n$ , assume  $x_i(t)$  to be continuous on  $I$  except possibly at partition points  $t_b, b = 1, \dots, p-1$ , where finite left and right limits exist; denote these limits by  $x_i(t_b^-)$  and  $x_i(t_b^+)$ , respectively. The amount of discontinuity of each member of  $x(t)$  at each partition point will be assumed known, and we write

$$x_i(t_b^+) - x_i(t_b^-) - d_{ib} = 0,$$

with each  $d_{ib}$  a known constant. Also we let  $x_i(t_b) = x_i(t_b^+)$ . Thus  $x_i(t)$  is continuous at  $t_b$  if and only if  $d_{ib} = 0$ .

Let  $y(t)$  denote the set  $(y_1(t), \dots, y_m(t))$ , where  $y_j(t)$  is piecewise continuous on  $I, j = 1, \dots, m$ , finite discontinuities being allowed between, as well as at, partition points. In the formulation of the problem the  $y_j(t)$  will occur only as undifferentiated variables and will not occur in the function to be minimized nor in the end and intermediate point constraints. Such variables are called control variables, while the  $x_i(t)$  are called state variables.

Differentiation with respect to  $t$  will be indicated by a superposed dot and partial derivatives by subscript variables. Each subscript or superscript index will always have the range specified when first used (and given in the table of notations), and repeated indices in a product will indicate summation unless the contrary is stated.

The problem will be to find in a class of admissible arcs

$$x(t), \quad y(t), \quad (t_0, \dots, t_p), \quad t_0 \leq t \leq t_p,$$

which satisfy differential equations

$$\dot{x}_i = L_i^a(t, x, y), \quad t \text{ in } I_a, \quad a = 1, \dots, p, \quad i = 1, \dots, n,$$

finite equations

$$M_g^a(t, x, y) = 0, \quad g = 1, \dots, q,$$

inequalities

$$N_h^a(t, x, y) \geq 0, \quad h = 1, \dots, r, \quad q + r \leq m,$$

and end and intermediate point conditions

$$J_k(t_0, \dots, t_p, x(t_0), x(t_1^-), x(t_1^+), \dots, x(t_p)) = 0,$$

$$k = 1, \dots, s < (n + 1)(p + 1),$$

$$x_i(t_b^+) - x_i(t_b^-) - d_{ib} = 0, \quad b = 1, \dots, p - 1,$$

one that will minimize

$$J_0(t_0, \dots, t_p, x(t_0), x(t_1^-), x(t_1^+), \dots, x(t_p)).$$

In order to state precisely the properties of the functions involved in the problem, let  $R_a$  be an open connected set in the  $m+n+1$  dimensional  $(t,x,y)$  space whose projection on the  $t$ -axis contains the interval  $I_a$ , and let  $S$  be an open connected set in the  $2np + p + 1$  dimensional space of points

$$(t_0, \dots, t_p, x(t_0), x(t_1^-), x(t_1^+), \dots, x(t_p)).$$

The functions  $L_i^a, M_g^a, N_h^a$  are assumed continuous with continuous partial derivatives through those of third order in  $R_a$ , and  $J_0, J_k$  are to have such continuity properties in  $S$ . For each  $a$ , the matrix

$$\begin{vmatrix} M_{gy_j}^a & 0 \\ N_{hy_j}^a & D_1^a \end{vmatrix}$$

is assumed of rank  $q + r$  in  $R_a$ , where  $D_1^a$  is an  $r$  by  $r$  diagonal matrix with  $N_1^a, \dots, N_r^a$  as diagonal elements. The matrix

$$\begin{vmatrix} J_{ct_0} & J_{ct_b} & J_{ct_p} & J_{cx_i(t_0)} & J_{cx_i(t_b^-)} & J_{cx_i(t_b^+)} & J_{cx_i(t_p)} \end{vmatrix}, c=0, \dots, s,$$

is assumed of rank  $s + 1$  in  $S$ .

An admissible set is a set  $(t,x,y)$  in  $R_a$  for some  $a = 1, \dots, p$ . An admissible sub-arc  $C_a$  is a set of functions  $x(t), y(t), t$  on  $I_a$ , with each  $(t,x,y)$  admissible, and such that  $x(t)$  is continuous and  $\dot{x}(t), y(t)$  are piecewise continuous on  $I_a$ . An admissible arc is a partition set  $(t_0, \dots, t_p)$  together with a set

of admissible sub-arcs  $C_a$ ,  $a = 1, \dots, p$ , such that the set  $(t_0, \dots, t_p, x(t_0), x(t_1^-), x(t_1^+), \dots, x(t_p))$  is in  $S$ .

### THE MULTIPLIER RULE

An admissible arc  $E$  for which

$$J_k(t_0, \dots, t_p, x(t_0), x(t_1^-), x(t_1^+), \dots, x(t_p)) = 0,$$

$$x_i(t_b^+) - x_i(t_b^-) - d_{ib} = 0,$$

is said to satisfy the multiplier rule if there exists a function

$$H(t, x, y, \lambda, \mu, \nu) = \lambda_i L_i^a - \mu_g M_g^a + \nu_h N_h^a,$$

with multipliers  $\lambda_i(t), \mu_g(t), \nu_h(t)$  continuous except possibly at partition points or corners of  $E$ , where finite left and right limits exist, such that for each  $t$  in  $I_a$ ,  $a = 1, \dots, p$ ,

$$(1) \quad \lambda_i = - \int_{t_{a-1}}^t H_{x_i} dt + c_i^a, \quad H_{y_j} = 0, \quad \dot{x}_i = L_i^a, \quad M_g^a = 0, \quad N_h^a \geq 0,$$

and such that the transversality matrix

$$(2) \quad \left\| \begin{array}{cccccc} H(t_0) & H(t_b^+) - H(t_b^-) & -H(t_p) & -\lambda_1(t_0) & -\lambda_1(t_b^+) + \lambda_1(t_b^-) & \lambda_1(t_p) \\ J_{ct_0} & J_{ct_b} & J_{ct_p} & J_{cx_i(t_0)} & J_{cx_i(t_b^+)} + J_{cx_i(t_b^-)} & J_{cx_i(t_p)} \end{array} \right\|$$

is of rank  $s + 1$ . The multipliers  $\nu_h$  are zero when  $N_h > 0$ . Every minimizing arc  $E$  must satisfy the multiplier rule.

It may be noted that the constants  $c_i^a$  are the initial values  $\lambda_i(t_{a-1}^+)$ , respectively, of the multipliers  $\lambda_i$  for the several stages.

Corollary. Between corners of a minimizing arc E the equations

$$\dot{x}_i = H_{\lambda_i}, \quad \dot{\lambda}_i = -H_{x_i}, \quad H_{y_j} = 0, \quad H_{\mu_g} = 0, \quad \nu_h H_{\nu_h} = 0 \quad (\text{not summed})$$

hold and hence also

$$\frac{dH}{dt} = H_t.$$

To prove the multiplier rule we transform the problem into a Denbow problem of the type treated by Hunt and summarized in the Appendix. Let  $z(t)$  denote the set  $(z_1(t), \dots, z_N(t))$ ,  $N = n + m + r$ , where

$$z_i(t) = x_i(t), \quad z_{n+j}(t) = \int_{t_0}^t y_j(t) dt, \quad z_{n+m+h}(t) = \int_{t_0}^t \sqrt{N_h(t, x(t), y(t))} dt;$$

or, equivalently,

$$(3) \quad \dot{z}_i(t) = x_i(t), \quad \dot{z}_{n+j}(t) = y_j(t), \quad \dot{z}_{n+m+h}(t) = \sqrt{N_h(t, x(t), y(t))}$$

with initial conditions  $z_{n+j}(t_0) = z_{n+m+h}(t_0) = 0$ . Note that for admissible arcs  $z_{n+j}$  and  $z_{n+m+h}$  are continuous on  $I$ . In the definition of  $\dot{z}_{n+m+h}$  the superscript  $a$  has been omitted from  $N_h$ . Where this is done, it is to be understood that  $N_h(t, x, y) = N_h^a(t, x, y)$  when  $t$  is in  $I_a$ . Similar usage applies to  $L_i$  and  $M_g$ .

Denote the differential equations for the transformed problem by

$$(4) \quad \phi_\beta^a(t, z, \dot{z}) = 0, \quad \beta = 1, \dots, n + q + r = M, \quad t \text{ in } I_a,$$

where

$$\begin{aligned}\phi_i^a &= \dot{z}_i - L_i^a(t, z_1, \dots, z_n, \dot{z}_{n+1}, \dots, \dot{z}_{n+m}) \\ (5) \quad \phi_{n+g}^a &= M_g^a(t, z_1, \dots, z_n, \dot{z}_{n+1}, \dots, \dot{z}_{n+m}) \\ \phi_{n+q+h}^a &= \dot{z}_{n+m+h}^2 - N_h^a(t, z_1, \dots, z_n, \dot{z}_{n+1}, \dots, \dot{z}_{n+m}).\end{aligned}$$

Let the conditions on end and intermediate points be denoted by

$$(6) \quad f_{\gamma}(t_0, \dots, t_p, z(t_0), z(t_1^-), z(t_1^+), \dots, z(t_p)) = 0, \quad \gamma = 1, \dots, s+m+r=K,$$

where

$$\begin{aligned}f_k &= J_k(t_0, \dots, t_p, z_1(t_0), \dots, z_n(t_0), z_1(t_1^-), \dots, z_n(t_1^-), z_1(t_1^+), \dots, \\ &\quad z_n(t_1^+), \dots, z_1(t_p), \dots, z_n(t_p)),\end{aligned}$$

$$f_{s+j} = z_{n+j}(t_0),$$

$$f_{s+m+h} = z_{n+m+h}(t_0),$$

plus the following difference relations at intermediate points,

$$(7) \quad z_{\alpha}(t_b^+) - z_{\alpha}(t_b^-) - d_{\alpha b} = 0, \quad \alpha = 1, \dots, N.$$

Note that  $d_{\alpha b} = 0$  for  $\alpha = n+1, \dots, N$ , since  $z_{n+j}$  and  $z_{n+m+h}$  are continuous. Let the transform of the function  $J_0$  be denoted by  $f_0$ .

Each point  $(t, x_1, \dots, x_n, y_1, \dots, y_m)$  of  $R_a$  transforms into a point  $(t, z_1, \dots, z_n, \dot{z}_{n+1}, \dots, \dot{z}_{n+m})$ . Let  $R'_a$  denote the open set in  $(2N+1)$ -dimensional  $(t, z, \dot{z})$  space whose restriction to the coordinates

$(t, z_1, \dots, z_n, \dot{z}_{n+1}, \dots, \dot{z}_{n+m})$  is the transform of  $R_a$ , the other coordinates of  $R'_a$  having unlimited range,  $-\infty$  to  $+\infty$ .

Let  $S'$  denote an open set in the  $2Np + p + 1$  dimensional space of points  $(t_0, \dots, t_1, z(t_0), z(t_1^-), z(t_1^+), \dots, z(t_p))$  whose restriction to  $(t_0, \dots, t_1, z_1(t_0), \dots, z_n(t_0), z_1(t_1^-), \dots, z_n(t_1^-), z_1(t_1^+), \dots, z_n(t_1^+), \dots, z_1(t_p), \dots, z_n(t_p))$  is the transform of  $S$  and which includes zero values for  $z_{n+1}(t_0), \dots, z_N(t_0)$ .

An admissible set for the transformed problem will be a set  $(t, z, \dot{z})$  in  $R'_a$  for some  $a$ . An admissible sub-arc  $C'_a$  will be a set of functions  $z(t)$  on  $I_a$  having each  $(t, z, \dot{z})$  admissible, with  $z(t)$  continuous and  $\dot{z}(t)$  piecewise continuous on  $I_a$ . An admissible arc will be a partition set  $(t_0, \dots, t_p)$  together with a set of admissible sub-arcs  $C'_a$  whose end and intermediate points lie in  $S'$ .

It follows from the assumptions about the functions  $L_i^a, M_g^a, N_h^a$  in  $R_a$  and  $J_o, J_k$  in  $S$  that the functions  $\phi_\beta^a$  and  $f_o, f_\gamma$  will have continuous partial derivatives to the third order in regions  $R'_a$  and  $S'$ , respectively. The matrix  $\|\phi_{\beta \dot{z}}^a\|$  can be readily verified to be of rank  $n + m + r$  for  $t$  in  $I_a$  since it can be written

$$\left\| \begin{array}{ccc} I & -L_{iyj}^a & 0 \\ 0 & M_{gyj}^a & 0 \\ 0 & -N_{hyj}^a & D_2^a \end{array} \right\|$$

where  $I$  is an  $n$  by  $n$  identity matrix and  $D_2^a$  is an  $r$  by  $r$  diagonal



matrix with diagonal elements  $2\dot{z}_{n+m+1}, \dots, 2\dot{z}_N$ . Also the matrix

$$\left\| \begin{matrix} f_{\rho t_0} & f_{\rho t_b} & f_{\rho t_p} & f_{\rho z(t_0)} & f_{\rho z(t_b^-)} & f_{\rho z(t_b^+)} & f_{\rho z(t_p)} \end{matrix} \right\|, \quad \rho = 0, 1, \dots, K,$$

is found to be of rank  $K + 1$ .

The assumptions made in the formulation of the problem in the Appendix are thus established, and hence the theorems of the Appendix can be applied. From equations (5) the required function  $F$  becomes

$$F(t, z, \dot{z}, \lambda, \mu, \nu) = \lambda_i (\dot{z}_i - L_i^a) + \mu_g M_g^a + \nu_h (\dot{z}_{n+m+h}^2 - N_h^a),$$

where the arguments of  $L_i^a$ ,  $M_g^a$ ,  $N_h^a$  are

$$(t, z_1, \dots, z_n, \dot{z}_{n+1}, \dots, \dot{z}_{n+m}), \quad t \text{ in } I_a, \quad a = 1, \dots, p,$$

and the multipliers  $\lambda_i(t)$ ,  $\mu_g(t)$ ,  $\nu_h(t)$  are continuous except possibly at corners or partition points, at which right and left limits exist.

Now define a function  $H$  whose arguments are

$(t, z_1, \dots, z_n, \dot{z}_{n+1}, \dots, \dot{z}_{n+m}, \lambda, \mu, \nu)$  as follows:

$$H = \lambda_i L_i^a - \mu_g M_g^a + \nu_h N_h^a.$$

The relationship between  $F$  and  $H$  is given by the equation

$$F = \lambda_i \dot{z}_i + \nu_h \dot{z}_{n+m+h}^2 - H,$$

and from equations (A-5) of the Appendix

$$\lambda_i = - \int_{t_{a-1}}^t H_{z_i} dt + c_i^a,$$

$$- H_{\dot{z}_{n+j}} = c_{n+j}^a,$$

$$- 2\gamma_h \dot{z}_{n+m+h} = c_{n+m+h}^a \quad (\text{not summed}).$$

Furthermore, the multiplier rule given in the Appendix establishes the existence of constants  $e_\rho$ , not all zero, satisfying the transversality conditions:

$$e_c J_{ct_0} + [\lambda_i \dot{z}_i + 2\gamma_h \dot{z}_{n+m+h}^2 - \dot{z}_{n+j} H_{\dot{z}_{n+j}}]^{t_0} = 0,$$

$$e_c J_{ct_b} + [\lambda_i \dot{z}_i + 2\gamma_h \dot{z}_{n+m+h}^2 - \dot{z}_{n+j} H_{\dot{z}_{n+j}}]_{t_b}^{t_b^+} = 0,$$

$$e_c J_{ct_p} + [\lambda_i \dot{z}_i + 2\gamma_h \dot{z}_{n+m+h}^2 - \dot{z}_{n+j} H_{\dot{z}_{n+j}}]_{t_p} = 0,$$

$$e_c J_{cz_i}(t_0) - [\lambda_i]^{t_0} = 0,$$

$$e_{s+j} - [-H_{\dot{z}_{n+j}}]^{t_0} = 0,$$

$$e_{s+m+h} - [2\gamma_h \dot{z}_{n+m+h}]^{t_0} = 0, \quad (\text{not summed}),$$

$$e_c (J_{cz_i}(t_b^+) + J_{cz_i}(t_b^-)) - [\lambda_i]_{t_b}^{t_b^+} = 0,$$

$$- [-H_{\dot{z}_{n+j}}]_{t_b}^{t_b^+} = 0,$$

$$- [2\gamma_h \dot{z}_{n+m+h}]_{t_b}^{t_b^+} = 0, \quad (\text{not summed}),$$

$$e_c^J c_{z_i}(t_p) - [\lambda_i]_{t_p} = 0,$$

$$- \left[ - H_{\dot{z}_{n+j}} \right]_{t_p} = 0,$$

$$- \left[ 2\nu_h \dot{z}_{n+m+h} \right]_{t_p} = 0, \quad (\text{not summed}).$$

Recalling the equations  $- H_{\dot{z}_{n+j}} = c_{n+j}^a$  and observing from the foregoing equations that the  $H_{\dot{z}_{n+j}}$  are continuous at partition points and zero at  $t_p$ , we have  $c_{n+j}^a = 0$  for each  $a$  and  $j$ . Similarly,  $- 2\nu_h \dot{z}_{n+m+h} = c_{n+m+h}^a$  (not summed), and the  $- 2\nu_h \dot{z}_{n+m+h}$  are continuous at partition points and zero at  $t_p$ ; hence  $c_{n+m+h}^a = 0$  for each  $a$  and  $h$ . Since  $\dot{z}_{n+m+h}^2 = N_h$ , this implies that  $\nu_h = 0$  when  $N_h > 0$ . It now follows that

$$\lambda_i \dot{z}_i + 2\nu_h \dot{z}_{n+m+h}^2 - \dot{z}_{n+j} H_{\dot{z}_{n+j}} = H,$$

and the first  $p + 1$  transversality equations become

$$e_c^J c_{t_0} + [H]_{t_0} = 0,$$

$$e_c^J c_{t_b} + [H]_{t_b}^+ = 0,$$

$$e_c^J c_{t_b} + [H]_{t_p} = 0.$$

Thus we have  $(p+1)(n+m+r+1)$  transversality equations; but, since

$e_p = 0$  for  $p > s$  (i.e., the last  $m+r$  of the  $e$ 's are zero), these may be reduced to only  $(p+1)(n+1)$  transversality equations. Changing variables to those of the original problem and writing this reduced set of transversality equations in equivalent matrix form completes the proof of the multiplier rule.

An extremal is an admissible arc and set of multipliers satisfying equations (1) and such that its functions  $\dot{x}(t), y(t), \lambda(t), \mu(t), \nu(t)$  have continuous first derivatives except possibly at partition points, where finite left and right limits exist. An extremal, or sub-arc of an extremal, is called non-singular if the determinant

$$\begin{vmatrix} H_{y_j y_e} & M_{gy_j} & N_{hy_j} & 0 \\ M_{gy_e} & 0 & 0 & 0 \\ N_{hy_e} & 0 & 0 & A \\ 0 & 0 & A & B \end{vmatrix}$$

is different from zero along it,  $A$  and  $B$  being diagonal matrices with diagonal elements  $\sqrt{N_1}, \dots, \sqrt{N_r}$  and  $\nu_1, \dots, \nu_r$ , respectively.

To define normal arcs, let the transversality conditions be used in equation form involving constant multipliers  $e_0, \dots, e_s$ , as in the proof given for the multiplier rule. An admissible arc with a set of multipliers  $\lambda_i, \mu_g, \nu_h, e_c$  satisfying the multiplier rule is then called normal if  $e_0 = 1$ . For this value of  $e_0$  the multipliers are

unique. On putting  $e_0 = 1$  in the transversality equations the following equivalent matrix form is obtained.

For a normal minimizing arc the transversality matrix

$$\begin{vmatrix} H(t_0) + J_{ot_0} & H(t_b^+) - H(t_b^-) + J_{ot_b} & -H(t_p) + J_{ot_p} & -\lambda_1(t_0) + J_{ox_1}(t_0) \\ J_{kt_0} & J_{kt_b} & J_{kt_p} & J_{kx_1}(t_0) \\ -\lambda_1(t_b^+) + \lambda_1(t_b^-) + J_{ox_1}(t_b^+) + J_{ox_1}(t_b^-) & \lambda_1(t_p) + J_{ox_1}(t_p) & & \\ J_{kx_1}(t_b^+) + J_{kx_1}(t_b^-) & & J_{kx_1}(t_p) & \end{vmatrix}$$

is of rank  $s$ .

Since the matrix is of order  $s + 1$  by  $(n+1)(p+1)$ , the requirement that the rank be  $s$  imposes  $(n+1)(p+1) - s$  conditions. This is one more condition than is imposed by the multiplier rule as first stated, the condition there being sufficient to determine the multipliers only up to an arbitrary proportionality factor.

#### WEIERSTRASS CONDITION

The  $\mathcal{E}$ -function of the Appendix becomes, on using  $F$  as given by equation (8),

$$\begin{aligned} \mathcal{E} = & \lambda_1 \dot{z}_1 + \nu_h \dot{z}_{n+m+h}^2 - H(\dot{z}) - \lambda_1 \dot{z}_1 - \nu_h \dot{z}_{n+m+h}^2 + H(\dot{z}) \\ & - (\dot{z}_1 - \dot{z}_1) \lambda_1 + (\dot{z}_{n+j} - \dot{z}_{n+j}) H_{\dot{z}_{n+j}}(\dot{z}) - (\dot{z}_{n+m+h} - \dot{z}_{n+m+h}) 2\nu_h \dot{z}_{n+m+h}, \end{aligned}$$

where the complete set of arguments in  $H(\dot{z})$  is

$(z_1, \dots, z_n, \dot{z}_{n+1}, \dots, \dot{z}_{n+m}, \lambda_1, \dots, \lambda_n, \mu_1, \dots, \mu_m, \nu_1, \dots, \nu_r)$  and in  $H(\dot{Z})$ , the same except that  $\dot{z}_{n+1}, \dots, \dot{z}_{n+m}$  replace  $\dot{z}_{n+1}, \dots, \dot{z}_{n+m}$ . Since  $\dot{z}_{n+j} = y_j$  and  $\dot{z}_{n+m+h} = \sqrt{N_h}$ , it follows from the multiplier rule that along a minimizing arc  $H_{\dot{z}_{n+j}}(\dot{z})$ ,  $\nu_h \dot{z}_{n+m+h}^2$ , and  $\nu_h \dot{z}_{n+m+h}$  are all zero. Hence, after simplification of  $\mathcal{E}$ , the Weierstrass condition is that for a normal minimizing arc  $E'$  the inequality

$$\mathcal{E} = H(\dot{z}) - H(\dot{Z}) + \nu_h \dot{z}_{n+m+h}^2 \geq 0$$

must hold at each element  $(t, z, \dot{z}, \lambda, \mu, \nu)$  of  $E'$  for all admissible sets  $(t, z, \dot{Z})$  satisfying  $M_g(t, z, \dot{Z}) = 0$  and  $\dot{Z}_{n+m+h}^2 - N_h(t, z, \dot{Z}) = 0$ . Let  $z_i$  be replaced by  $x_i$ ,  $\dot{z}_{n+j}$  by  $y_j$ ,  $\dot{Z}_{n+j}$  by  $Y_j$ , and  $\dot{Z}_{n+m+h}^2$  by  $N_h(t, x, Y)$ . Then, on referring to the definition of  $H$  and utilizing the facts that along a minimizing arc  $M_g(t, x, y) = 0$ ,  $\nu_h N_h(t, x, y) = 0$  (not summed) and that  $M_g(t, x, Y)$  is required to be zero in the Weierstrass condition, one can reduce the condition to the following form.

Weierstrass Condition. For a normal minimizing arc  $E$  the inequality

$$\lambda_i L_i(t, x, y) \geq \lambda_i L_i(t, x, Y)$$

must hold at each element  $(t, x, y, \lambda, \mu, \nu)$  of  $E$  for all admissible sets  $(t, x, Y)$  satisfying  $M_g(t, x, Y) = 0$  and  $N_h(t, x, Y) \geq 0$ .

#### CLEBSCH CONDITION

To apply the Clebsch condition of the Appendix to our transformed

problem we need the second partial derivatives of  $F$ . From equation (8) these are found to be

$$F_{\dot{z}_i \dot{z}_\alpha} = 0, \quad F_{\dot{z}_{n+j} \dot{z}_{n+e}} = -H_{\dot{z}_{n+j} \dot{z}_{n+e}}, \quad F_{\dot{z}_{n+j} \dot{z}_{n+m+h}} = 0,$$

$$F_{\dot{z}_{n+m+h} \dot{z}_{n+m+d}} = \begin{cases} 2\nu_h & \text{if } d = h, \\ 0 & \text{if } d \neq h. \end{cases}$$

On dropping the terms in the Clebsch inequality having zero coefficients, re-numbering the subscripts of the  $\pi$ 's in the remaining terms and denoting the last  $r$  of them by  $\theta_1, \dots, \theta_r$ , we can state that for a normal minimizing arc the inequality

$$-H_{\dot{z}_{n+j} \dot{z}_{n+e}} \pi_j \pi_e + 2\nu_h \theta_h^2 \geq 0$$

must hold at each element of  $E'$  for all  $\pi_1, \dots, \pi_m, \theta_1, \dots, \theta_r$  satisfying the equations  $M_{g\dot{z}_{n+j}} \pi_j = 0$ ,  $N_{h\dot{z}_{n+j}} \pi_j - 2\dot{z}_{n+m+h} \theta_h = 0$  ( $h$  not summed).

By the multiplier rule,  $\nu_h = 0$  at an element where  $N_h > 0$ . At an element where  $N_h = 0$ , and hence  $\dot{z}_{n+m+h} = 0$ , one may choose  $\theta_h \neq 0$  but  $\pi_1, \dots, \pi_m$  and the remaining  $\theta$ 's all zero. The Clebsch condition would then imply  $\nu_h \geq 0$ . Thus, for a normal minimizing arc the multipliers  $\nu_h$  are all non-negative.

Since  $\nu_h = 0$  when  $N_h > 0$ , it follows that at elements of a minimizing arc where  $N_h > 0$  the term  $2\nu_h \theta_h^2$  of the Clebsch inequality would drop out. When  $N_h = 0$  the term can also be dropped, for,

$\dot{z}_{n+m+h}$  would then be zero, and the condition  $N_{h\dot{z}_{n+j}} \pi_j - 2\dot{z}_{n+m+h} \theta_h = 0$  (h not summed) would be satisfied for any  $\theta_h$ . In particular the Clebsch condition would have to be satisfied with  $\theta_h = 0$  provided  $N_{h\dot{z}_{n+j}} \pi_j = 0$  and  $M_{g\dot{z}_{n+j}} \pi_j = 0$ . Thus the condition can finally be stated in the following form.

Clebsch Condition. For a normal minimizing arc E the inequality

$$H_{y_j y_e} \pi_j \pi_e \leq 0$$

must hold at each element  $(t, x, y, \lambda, \mu, \nu)$  of E for all sets  $\pi_1, \dots, \pi_m$  satisfying  $M_{gy_j}(t, x, y) \pi_j = 0$  and  $N_{hy_j}(t, x, y) \pi_j = 0$ , where in the last equation h ranges only over the subset of  $1, \dots, r$  for which  $N_h(t, x, y) = 0$ .



## APPENDIX

This appendix gives the formulation of the Denbow problem as modified by Hunt together with the multiplier rule and the necessary conditions analogous to those of Weierstrass and Clebsch. At the expense of some repetition, we have made this appendix independent of the main part of the paper.

Let  $t$  be the independent variable. For fixed  $p$ , define a set of variables  $(t_0, t_1, \dots, t_p)$  to be a partition set if and only if  $t_0 < t_1 < \dots < t_p$ . Let  $I$  denote the interval  $t_0 \leq t \leq t_p$  and  $I_a$  the subinterval  $t_{a-1} \leq t < t_a$  for  $a = 1, \dots, p-1$  and  $t_{a-1} \leq t \leq t_a$  for  $a = p$ . Let  $z(t)$  denote the set of functions  $(z_1(t), \dots, z_N(t))$ , where each  $z_\alpha(t)$ ,  $\alpha = 1, \dots, N$ , is continuous on  $I$  except possibly at partition points  $t_1, \dots, t_{p-1}$ . At these points right and left limits  $z_\alpha(t_1^-)$ ,  $z_\alpha(t_1^+)$ ,  $\dots$ ,  $z_\alpha(t_{p-1}^+)$  are assumed to exist and we let  $z_\alpha(t_b) = z_\alpha(t_b^+)$ ,  $b = 1, \dots, p-1$ .

The problem will be to find in a class of admissible arcs

$$z(t), \quad (t_0, \dots, t_p), \quad t_0 \leq t \leq t_p,$$

satisfying differential equations

$$(A-1) \quad \phi_\beta^a(t, z, \dot{z}) = 0, \quad t \text{ in } I_a, \quad \beta = 1, \dots, M < N,$$

and end and intermediate point conditions

$$(A-2) \quad f_\gamma(t_0, \dots, t_p, z(t_0), z(t_1^-), z(t_1^+), \dots, z(t_p)) = 0,$$

$$\gamma = 1, \dots, K \leq (N+1)(p+1),$$

$$(A-3) \quad z_{\alpha}(t_b^+) - z_{\alpha}(t_b^-) - d_{\alpha b} = 0$$

one that will minimize

$$f_0(t_0, \dots, t_p, z(t_0), z(t_1^-), z(t_1^+), \dots, z(t_p)).$$

Let  $R'_a$  be an open connected set in the  $2N+1$  dimensional  $(t, z, \dot{z})$  space whose projection on the  $t$ -axis contains  $I_a$ . The functions  $\phi_{\beta}^a$  are required to have continuous third partial derivatives in  $R'_a$  and each matrix  $\|\phi_{\beta}^a z_{\alpha}\|$  is assumed of rank  $M$  in  $R'_a$ . Let  $S'$  denote an open connected set in the  $2Np+p+1$  dimensional space of points  $(t_0, \dots, t_p, z(t_0), z(t_1^-), z(t_1^+), \dots, z(t_p))$  in which the functions  $f_{\rho}$ ,  $\rho = 0, 1, \dots, K$  have continuous third partial derivatives and the matrix

$$(A-4) \quad \left\| \begin{array}{ccccccc} f_{\rho t_0} & f_{\rho t_b} & f_{\rho t_p} & f_{\rho z_{\alpha}(t_0)} & f_{\rho z_{\alpha}(t_b^-)} & f_{\rho z_{\alpha}(t_b^+)} & f_{\rho z_{\alpha}(t_p)} \end{array} \right\|$$

is of rank  $K+1$ .

An admissible set is a set  $(t, z, \dot{z})$  in  $R'_a$  for some  $a=1, \dots, p$ . An admissible subarc  $C'_a$  is a set of functions  $z(t)$ ,  $t$  on  $I_a$ , with each  $(t, z, \dot{z})$  an admissible set and such that  $z(t)$  is continuous and  $\dot{z}(t)$  is piecewise continuous on  $I_a$ . An admissible arc  $E'$  is a partition set  $(t_0, \dots, t_p)$  together with a set of admissible subarcs  $C'_a$ ,  $a = 1, \dots, p$ , such that the set  $(t_0, \dots, t_p, z(t_0), z(t_1^-), z(t_1^+), \dots, z(t_p))$  is in  $S'$ .

Multiplier Rule. An admissible arc  $E'$  that satisfies equations (A-1), (A-2), (A-3) is said to satisfy the multiplier rule if there

exist constants  $e_\rho$  not all zero and a function

$$F(t, z, \dot{z}, \lambda) = \lambda_\rho \phi_\rho^a(t, z, \dot{z}), \quad t \text{ in } I_a,$$

with multipliers  $\lambda_\rho(t)$  continuous except possibly at corners or dis-  
continuities of  $E'$ , where left and right limits exist, such that the fol-  
lowing equations hold:

$$(A-5) \quad F_{\dot{z}_\alpha} = \int_{t_{a-1}}^t F_{z_\alpha} dt + c_\alpha^a, \quad t \text{ in } I_a,$$

$$e_\rho f_{\rho t_0} + \left[ \dot{z}_\alpha F_{\dot{z}_\alpha} \right]_{t_0}^{t_0} = 0,$$

$$e_\rho f_{\rho t_b} + \left[ \dot{z}_\alpha F_{\dot{z}_\alpha} \right]_{t_b^-}^{t_b^+} = 0,$$

$$e_\rho f_{\rho t_p} + \left[ \dot{z}_\alpha F_{\dot{z}_\alpha} \right]_{t_p} = 0,$$

$$e_\rho f_{\rho z_\alpha}(t_0) - \left[ F_{\dot{z}_\alpha} \right]_{t_0}^{t_0} = 0,$$

$$e_\rho (f_{\rho z_\alpha}(t_b^+) + f_{\rho z_\alpha}(t_b^-)) - \left[ F_{\dot{z}_\alpha} \right]_{t_b^-}^{t_b^+} = 0,$$

$$e_\rho f_{\rho z_\alpha}(t_p) - \left[ F_{\dot{z}_\alpha} \right]_{t_p} = 0.$$

Every minimizing arc must satisfy the multiplier rule.

An extremal is defined to be an admissible arc and set of multipliers

$$z_\alpha(t), (t_0, \dots, t_p), \lambda_\rho(t), \quad t_0 \leq t \leq t_p,$$

satisfying equations (A-1) and (A-5) and such that the functions

$\dot{z}_\alpha(t), \lambda_\beta(t)$  have continuous first derivatives except possibly at partition points, where finite left and right limits exist. An extremal is non-singular in case the determinant

$$\begin{vmatrix} F_{\dot{z}_\alpha \dot{z}_\eta} & \phi_{\delta \dot{z}_\alpha} \\ \phi_{\beta \dot{z}_\eta} & 0 \end{vmatrix} \quad \begin{matrix} \alpha, \eta = 1, \dots, N \\ \beta, \delta = 1, \dots, M \end{matrix}$$

is different from zero along it. An admissible arc with a set of multipliers satisfying the multiplier rule is called normal if  $e_0 = 1$ . With this value of  $e_0$  the set of multipliers is unique.

Weierstrass Condition. An admissible arc  $E'$  with a set of multipliers  $\lambda_\beta(t)$  is said to satisfy the Weierstrass condition if

$$\begin{aligned} \mathcal{E}(t, z, \dot{z}, \lambda, \dot{Z}) &= F(t, z, \dot{Z}, \lambda) - F(t, z, \dot{z}, \lambda) \\ &\quad - (\dot{Z}_\alpha' - \dot{z}_\alpha) F_{\dot{z}_\alpha}(t, z, \dot{z}, \lambda) \geq 0 \end{aligned}$$

holds at every element  $(t, z, \dot{z}, \lambda)$  of  $E'$  for all admissible sets  $(t, z, \dot{Z})$  satisfying the equations  $\phi_\beta^a = 0$ . Every normal minimizing arc must satisfy the Weierstrass condition.

Clebsch Condition. An admissible arc  $E'$  with a set of multipliers  $\lambda_\beta(t)$  is said to satisfy the Clebsch condition if

$$F_{\dot{z}_\alpha \dot{z}_\eta}(t, z, \dot{z}, \lambda) \pi_\alpha \pi_\eta \geq 0$$

holds at every element  $(t, z, \dot{z}, \lambda)$  of  $E'$  for all sets  $(\pi_1, \dots, \pi_N)$  satisfying the equations

$$\phi_{\beta \dot{z}_\alpha}^a(t, z, \dot{z}) \pi_\alpha = 0.$$

Every normal minimizing arc must satisfy the Clebsch condition.

#### REFERENCES

1. C. H. Denbow, "A Generalized Form of the Problem of Bolza", Contributions to the Calculus of Variations 1933-37, The University of Chicago Press, Chicago, 1937, pp. 449-484.
2. R. W. Hunt, "A Generalized Bolza-Mayer Problem with Discontinuous Solutions and Variable Intermediate Points", Given at Conference on Guidance and Space Flight Theory, Marshall Space Flight Center, Huntsville, Alabama, October 9-10, 1963.
3. G. A. Bliss, "Lectures on the Calculus of Variations", The University of Chicago Press, Chicago, 1946.
4. M. R. Hestenes, "A General Problem in the Calculus of Variations with Applications to Paths of Least Time", The Rand Corporation Research Memorandum RM-100, Santa Monica, California, March, 1950.
5. F. A. Valentine, "The Problem of Lagrange with Differential Inequalities as Added Side Conditions", Contributions to the Calculus of Variations 1933-37, The University of Chicago Press, Chicago, 1937, pp. 403-447.

AUBURN UNIVERSITY

N65 33052

A MAXIMUM PRINCIPLE RE-ENTRY STUDY

By

Joe W. Reece  
Grady R. Harmon

AUBURN, ALABAMA

A large, stylized handwritten mark, possibly a signature or initials, consisting of a large '2' shape with a horizontal line extending to the left.

# A MAXIMUM PRINCIPLE RE-ENTRY STUDY

By

Joe W. Reece  
Grady R. Harmon

Auburn University  
Auburn, Alabama

## SUMMARY

33052

The Maximum Principle of Pontryagin is used to find the point-to-point re-entry trajectory of a space vehicle with an offset center of gravity which will minimize the accumulated aerodynamic acceleration. The mathematical model used incorporates the yaw angle and the true angle of attack as control variables. The set of characteristic differential equations is written with both algebraic and differential constraints. A computation procedure is devised so that numerical solutions can be obtained on a digital computer.

*Author*

# LIST OF SYMBOLS

$G$	Gravitational constant
$m$	Mass of the vehicle
$M$	Mass of the earth
$\bar{X}$	Plumbline position vector
$\bar{x}_m$	Missile system position vector
$\bar{x}_a$	Aerodynamic system position vector
$ R $	Absolute value of the plumbline position vector
$R_o$	Earth's radius
$\phi_r$	Roll angle
$\phi_y$	Yaw angle
$\phi_p$	Pitch angle
SR	Sine $\phi_r$
CR	Cosine $\phi_r$
SY	Sine $\phi_y$
CY	Cosine $\phi_y$
SP	Sine $\phi_p$
CP	Cosine $\phi_p$
$\bar{F}_a$	Aerodynamic force in the aerodynamic coordinate system



$\bar{F}_{am}$	Aerodynamic force in the missile system
$\bar{F}_G$	Gravitational force in the plumblane system
A	Projected cross-sectional area of vehicle
q	Dynamic pressure
$f(\alpha, \alpha_y)$	Vehicle configuration function
$\bar{\omega}_E$	Earth's angular velocity vector in plumblane system
t	Time
$\bar{V}_R$	Relative velocity vector (Plumblane System)
$\bar{V}_r$	Relative velocity vector (Aerodynamic System)
$\bar{V}_{rm}$	Relative velocity vector (Missile System)
$\bar{W}$	Velocity vector for abnormal air movement in plumblane system

## I. INTRODUCTION

In this paper an attempt is made to treat the optimum re-entry problem in a simplified dynamical manner. The condition for optimality is that the integral  $\int (\text{DRAG})^2 dt$  be a minimum for fixed end points. The first order differential equations of translational motion and the algebraic equations defining the relative velocity vector are the constraints. It is assumed that the attractive force of the earth and the aerodynamic drag are the only forces influencing the vehicle's motion. The vehicle has an offset center of gravity which aids maneuverability. The performance analysis is based on the Pontryagin fixed end point problem with dual control variables.

## II. STATEMENT OF THE PROBLEM

The problem herein presented is that of determining from a given class of allowable trajectories the best one yielding mission fulfillment.

A space vehicle is assumed to initiate a re-entry into the earth's atmosphere from some initial point above the earth's surface. The influencing forces are the gravitational force of the earth and the aerodynamic force created by atmospheric drag. The prediction of the vehicle's performance is based on the assumption that a control system is desired which will satisfy the following criteria:

1. Minimization of the accumulated g-forces on the vehicle's occupants.
2. Capability of making a point landing.

In mathematical form the first of these becomes the minimization of the integral of the square of the total aerodynamic acceleration. The second can be accomplished by the proper choice of the initial auxiliary variables.

The performance problem thus formulated becomes the Pontryagin fixed end point problem, where the functional to be minimized has as constraints the first order equations of motion and the finite relative velocity equations. The boundary conditions are the initial and terminal values of position and velocity. The yaw and true angles of attack are taken as control variables.

Additional assumptions made are as follows:

1. The earth is a rotating sphere and the inverse gravity law holds.
2. The mass of the vehicle is invariant with respect to time.
3. The vehicle has an offset center of gravity which is invariant with respect to the vehicle.
4. The center of pressure is invariant with respect to the center of gravity.

### III. COORDINATE SYSTEMS

Three rectangular cartesian coordinate systems will be used in this paper. They are:

1. The plumblane space fixed coordinate system
2. The vehicle fixed missile system
3. The aerodynamic system.

#### A. PLUMBLANE SYSTEM

The plumblane system, Figure 1, has its origin at the earth's center with the Y axis parallel to the gravity gradient at the launch point. The X axis is parallel to the earth fixed launch azimuth and the Z axis is such as to form a right-handed system.

#### B. MISSILE SYSTEM

The missile system, Figure 1, is defined with its origin at the center of gravity of the vehicle and its  $y_m$  axis parallel to the longitudinal axis of the vehicle. The  $x_m$  and  $z_m$  axes are taken so as to form a right-handed system which is parallel to the plumblane system at the launch point.

As the vehicle moves along its trajectory, the missile system undergoes a displacement with respect to the plumblane system. In

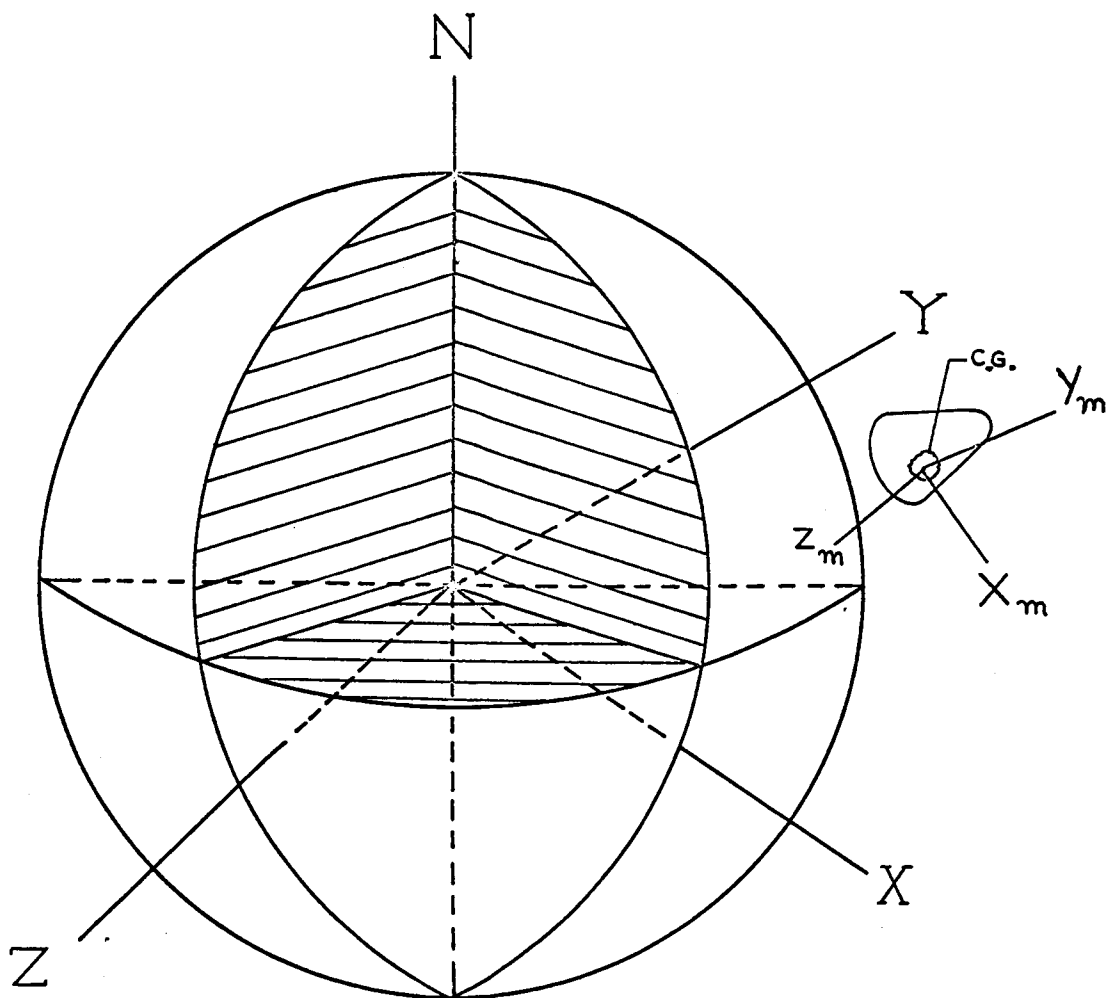


FIGURE 1.  
PLUMLINE AND MISSILE COORDINATE SYSTEMS .

flight the two coordinate systems are related through Eulerian angles which are measured by a gimbal. The flight direction of the vehicle is defined by first rotating about the Y axis by  $\phi_r$ , then about the new intermediate x axis by  $\phi_y$ , and finally about the z axis of the second intermediate system by  $\phi_p$ . All three rotations are considered positive counterclockwise when viewed from the positive end of the axis about which the rotation is taken (see Figure 2).

Thus, a position vector in the missile system may be written in terms of a position vector in the plumline system as

$$\bar{x}_m = [\phi_p] [\phi_y] [\phi_r] \bar{x}, \quad (1)$$

or

$$\begin{bmatrix} x_m \\ y_m \\ z_m \end{bmatrix} = \begin{bmatrix} CP & SP & 0 \\ -SP & CP & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & CY & SY \\ 0 & -SY & CY \end{bmatrix} \begin{bmatrix} CR & 0 & -SR \\ 0 & 1 & 0 \\ SR & 0 & CR \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (1a)$$

where CP designates cosine  $\phi_p$ , etc. Expanding the equation above gives

$$\bar{x}_m = \begin{bmatrix} CPCR + SPSYSR & SPCY & -CPSR + SPSYCR \\ -SPCR + CPSYSR & CPCY & SPSR + CPSYCR \\ CYSR & -SY & CYCR \end{bmatrix} \bar{x} = [A_D] \bar{x} \quad (1b)$$

### C. AERODYNAMIC SYSTEM

The aerodynamic system is defined with its origin at the center of pressure of the vehicle and its  $y_a$  axis coincident with the relative velocity vector. The  $x_a$  and  $z_a$  axes are chosen to form a right hand system.

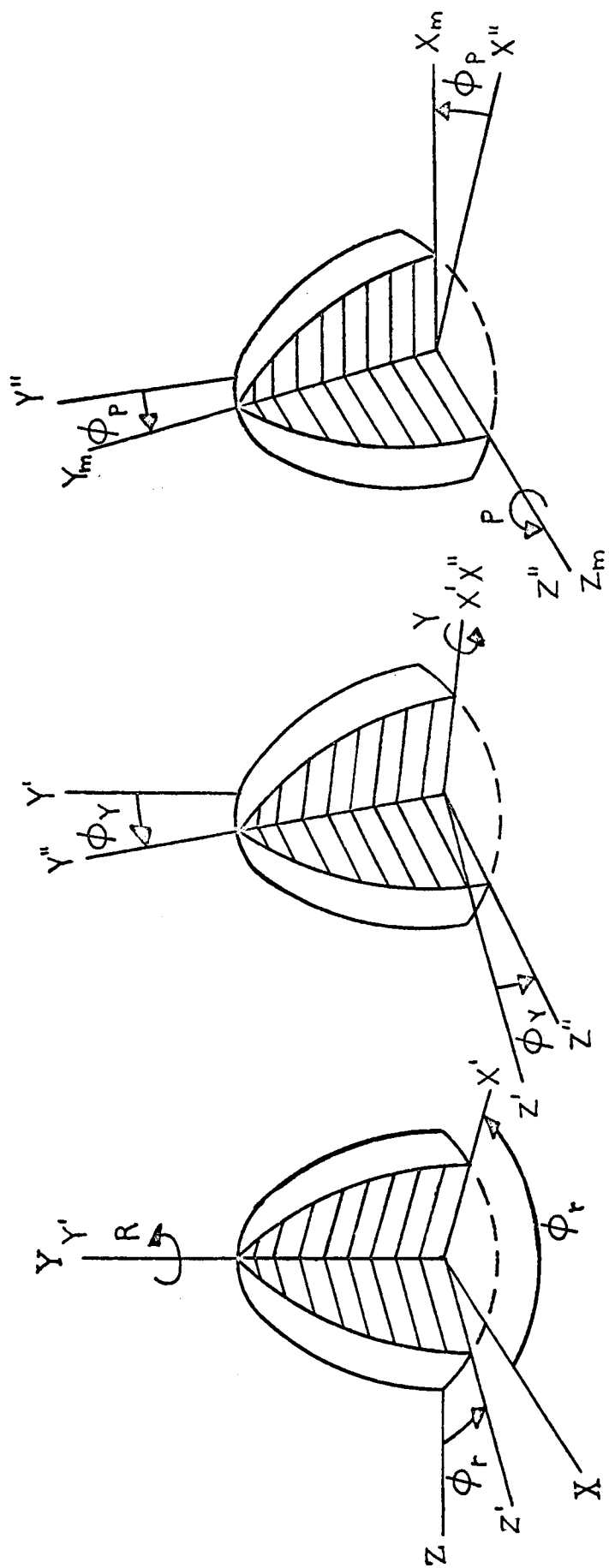


FIGURE 2. EULERIAN ANGLES



Again, as the vehicle moves in flight, there will be a displacement of the missile and aerodynamic coordinate systems relative to one another. The direction of the relative velocity vector or the  $y_a$  axis may be defined by the following rotations:

1. Rotate the vehicle fixed reference frame about the  $y_m$  axis such that the  $x_m$  axis is brought to lie in the plane which contains the  $y_m$  axis and the relative velocity vector. Denote this angle as  $\alpha_y$ .
2. Rotate about the new  $z$  axis to bring the  $y_m$  axis coincident with the relative velocity vector. Denote this angle as  $\alpha$ .

This angle is the so-called true angle of attack.

A position vector may now be written in the aerodynamic system in terms of the missile system as

$$\bar{x}_a = -[\alpha] [\alpha_y] \bar{x}_m, \quad (2)$$

or

$$\begin{bmatrix} x_a \\ y_a \\ z_a \end{bmatrix} = \begin{bmatrix} C\alpha & -S\alpha & 0 \\ S\alpha & C\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} C\alpha_y & 0 & -S\alpha_y \\ 0 & 1 & 0 \\ S\alpha_y & 0 & C\alpha_y \end{bmatrix} \begin{bmatrix} x_m \\ y_m \\ z_m \end{bmatrix} \quad (2a)$$

$$\bar{x}_a = \left[ \begin{array}{cc|c|cc} C\alpha & C\alpha_y & -S\alpha & -C\alpha & S\alpha_y \\ C\alpha_y & S\alpha & C\alpha & -S\alpha & S\alpha_y \\ S\alpha_y & 0 & 0 & C\alpha_y & \end{array} \right] \bar{x}_m = [A_a] \bar{x}_m. \quad (2b)$$

Figure 3 illustrates this system.

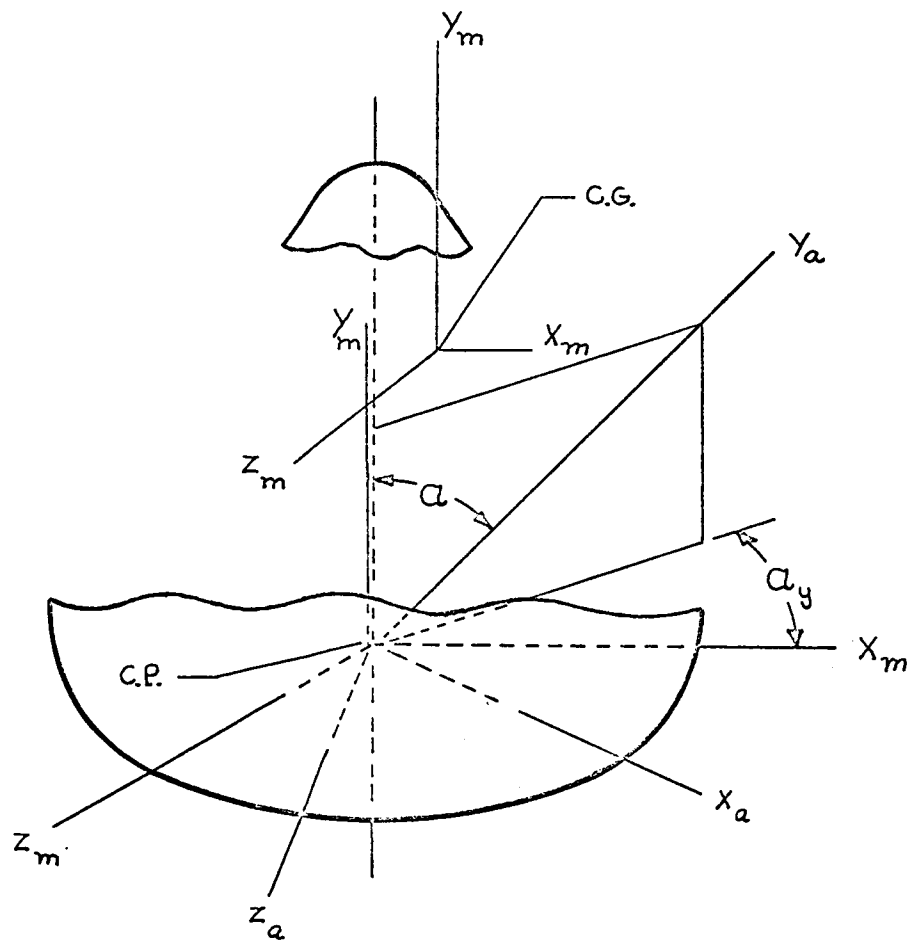


FIGURE 3. MISSILE AND AERODYNAMIC  
COORDINATE SYSTEMS

#### IV. BASIC MECHANICS

Gravitational force. Since a spherical earth was assumed, Newton's Law of Universal Gravitation which gives us an attractive force between the earth and the vehicle is

$$\bar{F}_G = - \frac{GMm\bar{X}}{|R|^3} \quad (3)$$

Aerodynamic force. The aerodynamic force, Figure 4, is a force due to atmospheric drag. It acts through the center of pressure and the direction of the force is always parallel and opposite to the relative velocity vector. Written in the aerodynamic system the force takes the following form:

$$\bar{F}_a = \begin{bmatrix} 0 \\ -F_a \\ 0 \end{bmatrix}. \quad (4)$$

In the missile system

$$\bar{F}_{am} = [A_a]^T \bar{F}_a, \quad (5)$$

or

$$\bar{F}_{am} = \begin{bmatrix} F_{amx} \\ F_{amy} \\ F_{amz} \end{bmatrix} = \begin{bmatrix} -F_a & S\alpha & C\alpha_y \\ -F_a & C\alpha & \\ F_a & S\alpha & S\alpha_y \end{bmatrix} \quad (5a)$$

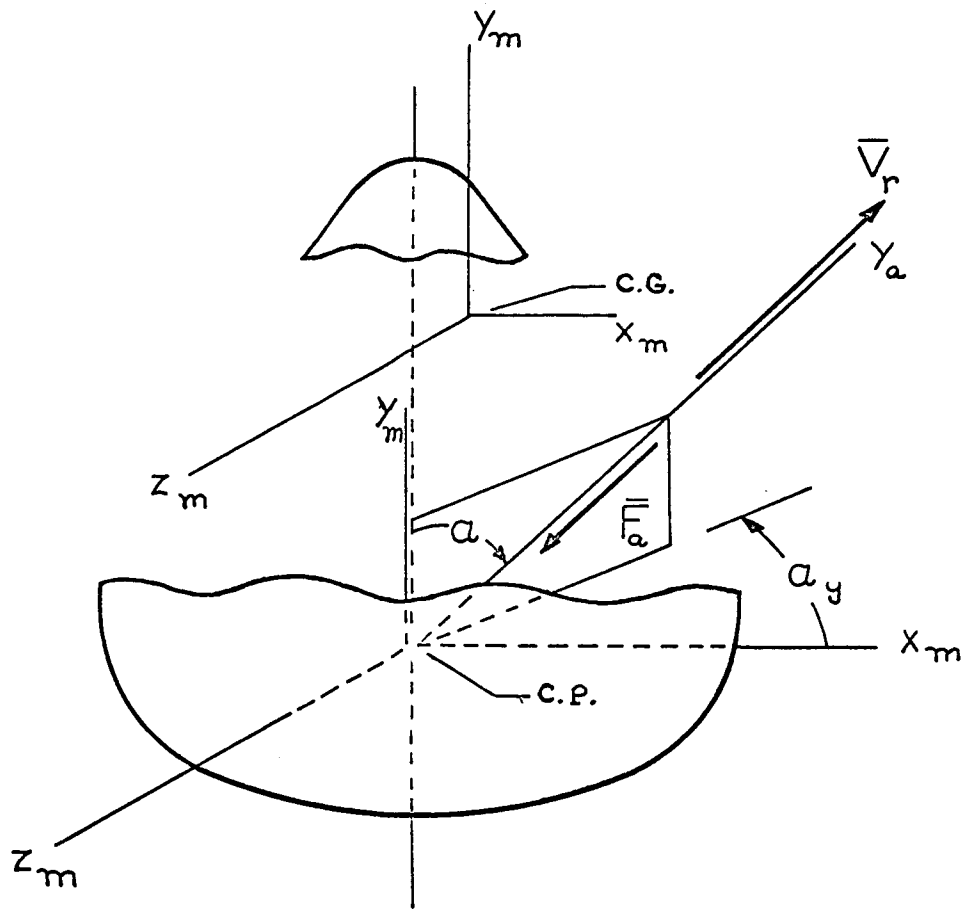


FIGURE 4. AERODYNAMIC FORCE SYSTEM

$F_a = Aqf(\alpha, \alpha_y)$ .  $A$  is the projected cross-section area of the vehicle,  $q$  the dynamic pressure, and  $f(\alpha, \alpha_y)$  a factor which is determined by the vehicle's configuration.

Since the aerodynamic force is dependent upon the relative velocity or the flow of air over the missile, it is appropriate at this time to discuss this flow. It is assumed that the atmosphere in the large moves with the earth. This gives at all times an air mass movement with respect to the plumbline system of

$$\bar{X} \times \bar{\omega}_E - \bar{W},$$

where  $\bar{W}$  is used to represent any abnormal air movement desired. The relative velocity vector in the plumbline system is then given by

$$\bar{V}_R = \dot{\bar{X}} + [\bar{X} \times \bar{\omega}_E - \bar{W}], \quad (6)$$

or

$$\begin{bmatrix} V_{RX} \\ V_{RY} \\ V_{RZ} \end{bmatrix} = \begin{bmatrix} \dot{X} \\ \dot{Y} \\ \dot{Z} \end{bmatrix} + \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \times \begin{bmatrix} \omega_{EX} \\ \omega_{EY} \\ \omega_{EZ} \end{bmatrix} - \begin{bmatrix} W_X \\ W_Y \\ W_Z \end{bmatrix}. \quad (6a)$$

In the missile system the relative velocity may be written as

$$\bar{V}_{rm} = [A_D] \bar{V}_R = \begin{bmatrix} V_{rmx} \\ V_{rmy} \\ V_{rmz} \end{bmatrix}, \quad (7)$$

or in terms of the aerodynamic system variables

$$\bar{V}_{rm} = [A_a]^T \bar{V}_r, \quad (8)$$

where

$$\bar{V}_r = \begin{bmatrix} 0 \\ V_r \\ 0 \end{bmatrix}$$

## V. EQUATIONS OF MOTION

As previously stated, only gravitational and aerodynamic forces are considered. Using Newton's Second Law, the translational motion of the center of gravity with respect to the plumbline system is given by the following set of second order differential equations.

$$\ddot{\bar{X}} = - \frac{GM\bar{X}}{|\bar{R}|^3} + \frac{[A_D]^T}{m} \bar{F}_{am}, \quad (9)$$

where

$$\ddot{\bar{X}} = \begin{bmatrix} \ddot{X} \\ \ddot{Y} \\ \ddot{Z} \end{bmatrix}$$

By making the following change of variable, the second order equations of translational motion may be reduced to first order.

$$\bar{u} = \begin{bmatrix} u \\ v \\ w \end{bmatrix} \equiv \begin{bmatrix} \dot{X} \\ \dot{Y} \\ \dot{Z} \end{bmatrix} = \dot{\bar{X}}. \quad (10)$$

The first order translational equations thus become

$$\dot{\bar{u}} = - \frac{GM\bar{X}}{|\bar{R}|^3} + \frac{[A_D]^T}{m} \bar{F}_{am}. \quad (11)$$

For convenience, the following definitions are made:

$$g \equiv - \frac{GM}{|R|^3}, \quad (12)$$

$$\frac{[A_D]^T}{m} \bar{F}_{am} \equiv \frac{F_a}{m} \bar{N} \equiv F'_a \bar{N}, \quad (13)$$

where

$$\bar{N} = \frac{[A_D]^T [A_a]^T \bar{F}_a}{F_a}$$

or

$$\bar{N} = \begin{bmatrix} - (CPCR + SPSYSR) C\alpha C\alpha_y + (SPCR - CPSYSR) C\alpha + CYSRSaSa_y \\ - SPCYC\alpha_y Sa - CPCYC\alpha - SYSaSa_y \\ (CPSR - SPSYCR) C\alpha_y Sa - (SPSR + CPSYCR) C\alpha + CYCRSaSa_y \end{bmatrix} \quad (14)$$

Thus, the translational equations may be written as

$$\ddot{\bar{u}} = F'_a \bar{N} + g\bar{X} \quad (15)$$



## VI. FORMULATION OF THE VARIATIONAL PROBLEM

The formulation of the variational problem requires further consideration of the constraint equations emanating from

$$\bar{V}_{rm} = [A_D] \bar{V}_R = [A_a]^T \bar{V}_r \quad (16)$$

or

$$\bar{V}_{rm} = [\phi_p] [\phi_y] \begin{bmatrix} a \\ b \\ c \end{bmatrix} \quad (17)$$

where

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = [\phi_r] \bar{V}_R \quad (18)$$

is the relative velocity vector referred to the intermediate system located by  $[\phi_r]$ . The system of equations (17) is solved for  $\phi_y$  and  $\phi_p$  to yield (see Appendix)

$$SP = \frac{V_{rmx}}{\sqrt{V_{rmx}^2 + V_{rmy}^2}}, \quad CP = \frac{V_{rmy}}{\sqrt{V_{rmx}^2 + V_{rmy}^2}} \quad (19)$$

$$-\pi \leq \phi_p \leq \pi$$

$$SY = \frac{-bV_{rmz} + c\sqrt{V_r^2 - V_{rmz}^2}}{V_r^2}, \quad CY = \frac{cV_{rmz} + b\sqrt{V_r^2 - V_{rmz}^2}}{V_r^2} \quad (20)$$

$$-\pi \leq \phi_y \leq \pi$$

The roll angle,  $\phi_r$ , is given by

$$SR = \frac{V_{RX}}{\sqrt{V_{RX}^2 + V_{RZ}^2}}, \quad CR = \frac{V_{RZ}}{\sqrt{V_{RX}^2 + V_{RZ}^2}} \quad (21)$$

$$-\pi \leq \phi_r \leq \pi$$

and is obtained by limiting  $\phi_y$  and  $\phi_p$  to real values. (See Appendix.)

As expressed in the problem statement, it is desired to determine from a given class of allowable trajectories the best one yielding mission fulfillment. This is accomplished by finding among all sets of admissible control  $\alpha(t)$ ,  $\alpha_y(t)$  which transfer the vehicle from  $\bar{X}_0$  to  $\bar{X}_T$  one for which the functional:

$$D = \int_{t_0}^{t_T} [DRAG]^2 dt \quad (22)$$

takes on a minimum value. In this analysis the word drag will be used synonymously with aerodynamic acceleration. Thus from Equation (15),

$$[DRAG]^2 = F'_a \bar{N} \cdot F'_a \bar{N} = (F'_a)^2 \bar{N} \cdot \bar{N} = (F'_a)^2, \quad (23)$$

and

$$D = \int_{t_0}^{t_T} (F'_a)^2 dt \quad ; \quad \dot{D} = (F'_a)^2. \quad (24)$$

The Pontryagin H function may now be written as follows:

$$H = \bar{\lambda}_I \cdot \dot{\bar{X}} + \bar{\lambda}_{II} \cdot \dot{\bar{u}} + \lambda_7 \dot{D}, \quad (25)$$

where

$$\bar{\lambda}_I = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} \text{ and } \bar{\lambda}_{II} = \begin{bmatrix} \lambda_4 \\ \lambda_5 \\ \lambda_6 \end{bmatrix}$$

The  $\lambda_i(t)$ ,  $i = 1 \dots 7$ , are the auxiliary variables that are incorporated in the same manner as the Lagrange multipliers in the classical calculus of variations. Substituting into H from Equation (15) results in the following:

$$H = \bar{\lambda}_I \cdot \bar{u} + \bar{\lambda}_{II} \cdot [F'_a \bar{N} + g \bar{X}] + \lambda_7 (F'_a)^2. \quad (26)$$

The expressions for the auxiliary variables are obtained from the H function and take the following form:

$$\begin{aligned} -\dot{\bar{\lambda}}_I = \frac{\partial H}{\partial \bar{X}} &= F'_a \frac{\partial (\bar{\lambda}_{II} \cdot \bar{N})}{\partial \bar{X}} + (\bar{\lambda}_{II} \cdot \bar{N}) \frac{\partial F'_a}{\partial \bar{X}} \\ &+ \bar{\lambda}_{II} g + (\bar{\lambda}_{II} \cdot \bar{X}) \frac{\partial g}{\partial \bar{X}} + \lambda_7 \frac{\partial (F'_a)^2}{\partial \bar{X}} \end{aligned} \quad (27)$$

$$\begin{aligned} -\dot{\bar{\lambda}}_{II} = \frac{\partial H}{\partial \bar{u}} &= \bar{\lambda}_I + F'_a \frac{\partial (\bar{\lambda}_{II} \cdot \bar{N})}{\partial \bar{u}} + (\bar{\lambda}_{II} \cdot \bar{N}) \frac{\partial F'_a}{\partial \bar{u}} \\ &+ \lambda_7 \frac{\partial (F'_a)^2}{\partial \bar{u}} \end{aligned} \quad (28)$$

$$-\dot{\lambda}_7 = \frac{\partial H}{\partial D} = 0 \quad (29)$$

It is implied from Equation (29) that  $\lambda_7 = \text{constant}$ . The equations to be solved for the control variables are given below.

$$\frac{\partial H}{\partial \alpha_y} = F'_a (\bar{\lambda}_{II} \cdot \frac{\partial \bar{N}}{\partial \alpha_y}) + (\bar{\lambda}_{II} \cdot \bar{N}) \frac{\partial F'_a}{\partial \alpha_y} \quad (30)$$

$$+ \lambda_7 \frac{\partial (F'_a)^2}{\partial \alpha_y} = 0$$

$$\frac{\partial H}{\partial \alpha} = F'_a (\bar{\lambda}_{II} \cdot \frac{\partial \bar{N}}{\partial \alpha}) + (\bar{\lambda}_{II} \cdot \bar{N}) \frac{\partial F'_a}{\partial \alpha} \quad (31)$$

$$+ \lambda_7 \frac{\partial (F'_a)^2}{\partial \alpha} = 0$$

Equations (15) and (19) through (21) are the constraint and definition equations which must be satisfied, and Equations (27) through (31) are the characteristic equations. The complete set of algebraic and differential equations needed for the problem solution have thus been found. The desired minimum drag re-entry path will thus be one which satisfies all the aforementioned equations. A closed form solution to this set of equations does not seem probable nor is the time spent in searching for such a solution justifiable since numerical solutions via digital computers can be achieved to almost any degree of accuracy.

## VII. COMPUTATIONAL PROCEDURE

### Functional Analysis:

When composing a computational procedure, it is sometimes found convenient to write the equations in functional form. Listed below is such a set.

$$\phi_r = \phi_r (\dot{\bar{x}})$$

$$\phi_y = \phi_y (\alpha, \alpha_y, \phi_r, \dot{\bar{x}})$$

$$\phi_p = \phi_p (\alpha, \alpha_y)$$

$$\dot{\bar{u}} = \dot{\bar{u}} (\bar{x}, \bar{u}, \bar{\phi}, \alpha, \alpha_y)$$

$$H = H (\bar{x}, \bar{u}, \bar{\phi}, \bar{\lambda}, \alpha, \alpha_y)$$

$$\dot{\bar{\lambda}} = \dot{\bar{\lambda}} (\bar{x}, \bar{u}, \bar{\phi}, \bar{\lambda}, \alpha, \alpha_y)$$

$$\frac{\partial H}{\partial \alpha} = \frac{\partial H}{\partial \alpha} (\bar{x}, \bar{u}, \bar{\phi}, \bar{\lambda}, \alpha, \alpha_y) = 0$$

$$\frac{\partial H}{\partial \alpha_y} = \frac{\partial H}{\partial \alpha_y} (\bar{x}, \bar{u}, \bar{\phi}, \bar{\lambda}, \alpha, \alpha_y) = 0$$

### Starting Values:

m

GM

R<sub>o</sub>

A

$\lambda_{70} = +1$

$$\bar{x}_0 = \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix}, \quad \bar{x}_{cp} = \begin{bmatrix} x_{cp} \\ y_{cp} \\ z_{cp} \end{bmatrix}, \quad \bar{u}_0 = \begin{bmatrix} u \\ v \\ w \end{bmatrix}$$

$$\bar{\lambda}_{IO} = \begin{bmatrix} \lambda_{10} \\ \lambda_{20} \\ \lambda_{30} \end{bmatrix}, \quad \bar{\lambda}_{II0} = \begin{bmatrix} \lambda_{40} \\ \lambda_{50} \\ \lambda_{60} \end{bmatrix}$$

Atmospheric tables for  $p$  as a function of altitude.

Atmospheric tables for  $\bar{W}$  as a function of position.

Aerodynamic tables for  $f(\alpha, \alpha_y)$  as a function of  $(\alpha, \alpha_y)$ .

"N" Line Computation:

- (1) Using starting values, iterate Equations (30) and (31) simultaneously for  $\alpha$  and  $\alpha_y$ .
- (2) Use  $(\alpha, \alpha_y)$  from Step (1) along with starting values to compute the following in order.

$\phi_r$  from Equation (21)

$\phi_y$  from Equation (20)

$\phi_p$  from Equation (19)

$\dot{\bar{u}}$  from Equation (15)

$H$  from Equation (26)

$\dot{\bar{\lambda}}_I$  from Equation (27)

$\dot{\bar{\lambda}}_{II}$  from Equation (28)

(3) Integrate to obtain the following:

$$\dot{\bar{u}} \text{ for } \bar{u} \text{ for } \bar{X}$$

$$\dot{\bar{\lambda}}_I \text{ for } \bar{\lambda}_I$$

$$\dot{\bar{\lambda}}_{II} \text{ for } \bar{\lambda}_{II}$$

(4) Use integrated values from Step (3) as starting values for the  $n + 1$  line.

Cut-off Criteria:

$$|\bar{V}_R| \leq \text{Mach } 2$$

## BIBLIOGRAPHY

- Bliss, G. A. Lectures on the Calculus of Variations. Chicago: The University of Chicago Press, 1946.
- Goldstein, Herbert. Classical Mechanics. Reading, Massachusetts: Addison-Wesley Publishing Company, Inc., 1959.
- Kopp, Richard E. Pontryagin Maximum Principle, Chapter 7 of Optimization Techniques. Edited by George Leitmann. Berkeley, California: Academic Press, 1961.
- Miner, W. E. Methods for Trajectory Computation, NASA-Marshall Space Flight Center, Internal Note, May 10, 1961.
- Pontryagin, L. S., et al. The Mathematical Theory of Optimal Processes. New York: Interscience Publishers, 1962.
- Progress Report No. 4 on Studies in the Fields of Space Flight and Guidance Theory, MTP-AERO-63-65. NASA-Marshall Space Flight Center, September 19, 1963.
- Progress Report No. 5 on Studies in the Fields of Space Flight and Guidance Theory. NASA-TMX-53024, March 17, 1964.



### VIII. CONCLUSION

The Maximum Principle has been employed to study the problem of minimizing the integral of drag squared.

The cut-off criterion on the trajectory was  $|V_R| \leq \text{Mach } 2$ . This is a reasonable criterion since the expression used for the aerodynamic force is valid only for velocity  $> \text{Mach } 2$ . If the desired terminal position is not attained simultaneously with the cut-off criterion, then a different set of initial auxiliary variables must be chosen. This procedure must continue until all of the terminal conditions are simultaneously satisfied.

No procedure has been developed in this paper for determining the initial auxiliary variables. An attempt is being made to formulate the transversality conditions for the problem and to apply the gradient method as an aid to numerical solution.

The problem, as formulated, is assured of a necessary but not sufficient condition for the existence of an optimum.

# APPENDIX

## SOLUTION FOR $\phi_r$ , $\phi_y$ , AND $\phi_p$

A first algebraic solution of the set of Equations (17) for  $\phi_p$  and  $\phi_y$  yields

$$SP = \frac{-a V_{rmy} + V_{rmx} \sqrt{V_{rmx}^2 + V_{rmy}^2 - a^2}}{V_{rmx}^2 + V_{rmy}^2} \quad (A1)$$

$$CP = \frac{a V_{rmx} + V_{rmy} \sqrt{V_{rmx}^2 + V_{rmy}^2 - a^2}}{V_{rmx}^2 + V_{rmy}^2} \quad (A2)$$

$$SY = \frac{-b V_{rmz} + c \sqrt{b^2 + c^2 - V_{rmz}^2}}{b^2 + c^2} \quad (A3)$$

$$CY = \frac{c V_{rmz} + b \sqrt{b^2 + c^2 - V_{rmz}^2}}{b^2 + c^2} \quad (A4)$$

First,  $\phi_p$  and  $\phi_y$  are limited to real values by setting  $a = 0$ .

It is easily shown that

$$V_{rmx}^2 + V_{rmy}^2 - a^2 = b^2 + c^2 - V_{rmz}^2.$$

The choice  $a = 0$  is allowable because of the dependency of the set of

Equations (17). As only two of the angles are required to locate a vector in three-space, no unique solution exists for  $\phi_r$ ,  $\phi_p$ , and  $\phi_y$ . However,  $a = 0$  provides that

$$\left. \begin{aligned} SR &= \frac{V_{RX}}{\sqrt{V_{RX}^2 + V_{RZ}^2}}, \quad CR = \frac{V_{RZ}}{\sqrt{V_{RX}^2 + V_{RZ}^2}} \\ -\pi &\leq \phi_r \leq \pi \end{aligned} \right\} \quad (A5)$$

Now, the corresponding values of  $\phi_p$  and  $\phi_y$  must be unique. To settle the choice of sign in Equations (A1) through (A4), it is recalled that the determinants of the right-handed rotation matrices  $[\phi_y]$  and  $[\phi_p]$  must be equal to unity. This eliminates two of the four possible sign combinations in Equations (A1) and (A2). The choice between the two remaining possibilities is made according to the relation between the aerodynamic coordinate system and the relative velocity vector  $\bar{V}_R$  set forth in Section IV. The resulting equations are

$$\left. \begin{aligned} SP &= \frac{V_{rmx}}{\sqrt{V_{rmx}^2 + V_{rmy}^2}}, \quad CP = \frac{V_{rmy}}{\sqrt{V_{rmx}^2 + V_{rmy}^2}} \\ -\pi &\leq \phi_p \leq \pi \end{aligned} \right\} \quad (A6)$$

and

$$\left. \begin{aligned}
 SY &= \frac{-b V_{rmz} + c \sqrt{b^2 + c^2 - V_{rmz}^2}}{b^2 + c^2} \\
 CY &= \frac{c V_{rmz} + b \sqrt{b^2 + c^2 - V_{rmz}^2}}{b^2 + c^2} \\
 -\pi &\leq \phi_y \leq \pi
 \end{aligned} \right\} (A7)$$

An additional result of setting  $a = 0$  can be seen from Equation (18) as

$$\left. \begin{aligned}
 b &= V_{RY} \\
 c &= \sqrt{V_{RX}^2 + V_{RZ}^2}
 \end{aligned} \right\} (A8)$$

Thus,  $b^2 + c^2 = V_R^2 = V_r^2$  and (A7) becomes

$$\left. \begin{aligned}
 SY &= \frac{-b V_{rmz} + c \sqrt{V_r^2 - V_{rmz}^2}}{V_r^2} \\
 CY &= \frac{c V_{rmz} + b \sqrt{V_r^2 - V_{rmz}^2}}{V_r^2} \\
 -\pi &\leq \phi_y \leq \pi
 \end{aligned} \right\} (A9)$$

The limits on  $\phi_r$ ,  $\phi_p$ , and  $\phi_y$  are chosen to provide a full revolution of freedom and eliminate any excess motion.

N65 33053

AN ENSEMBLE AVERAGING APPROACH TO  
OPTIMAL GUIDANCE POLYNOMIAL APPROXIMATIONS

Henry J. Kelley  
and  
Walter F. Denham

Contract NAS 8-11241

November 1964

Analytical Mechanics Associates, Inc.  
Uniondale, L.I., N.Y.

AN ENSEMBLE AVERAGING APPROACH TO  
OPTIMAL GUIDANCE POLYNOMIAL APPROXIMATIONS\*

Henry J. Kelley\*\* and Walter F. Denham\*\*\*  
Analytical Mechanics Associates, Inc.

ABSTRACT

33053

An approach to optimal guidance synthesis is developed in which an ensemble-averaged second order approximation to the performance function is minimized subject to constraints on the means and variances of other functions. The minimization is with respect to coefficients of assumed polynomial approximations of a linear feedback control law (in which the state is perfectly known) and coefficients in a linear termination law. A brief comparison is drawn with deterministic neighboring extremal control. While attention is directed mainly to first order necessary conditions, some comments are made on numerical solution by first and second order successive approximation methods. Extensions to include disturbances other than initial errors and to include state estimation errors are discussed briefly.

*Author*

---

\* This research was performed under Contract NAS 8-11241 with the Aero and Astrodynamics Division of NASA Marshall Space Flight Center, Huntsville, Alabama.

\*\* Vice-President

\*\*\* Senior Analyst

## Introduction

The earliest theoretical approaches to optimal guidance (Refs. 1 and 2) lead to computational methods for synthesizing linear feedback systems furnishing an approximation optimal to second order in an expansion about a given optimal reference trajectory. While the resulting systems fulfill their theoretical promise in providing high performance, terminal accuracy is found to be wanting, and the practical mechanization of the feedback law is encumbered by the need for storing time-varying "gains". Recent studies of the terminal accuracy problem (Refs. 3 and 4) indicate that a large improvement may be realized by transverse state comparison with the reference trajectory and suggest that this relatively simple procedure may be more effective than the addition of quadratic terms in the feedback approximation.

The present paper reports an idea for a synthesis scheme in which an ensemble-averaged second order approximation to the performance index is minimized with respect to certain parameters. These parameters include the coefficients in three polynomials in time which are used in place of general time-varying functions. Polynomial approximations are used for (1) the control programs of the optimal reference trajectory; (2) the state variable histories of the optimal reference trajectory; (3) the feedback gains for the assumed linear feedback control system. Additional parameters to be optimized are the coefficients in an assumed linear rule for termination of perturbed trajectories. The treatment is based upon the statistical methods pioneered in Refs. 5 and 6 in connection with synthesis of optimal midcourse guidance approximations.

## Formulation of the Problem

The dynamical system under consideration satisfies

$$\dot{x} = f(x, u, t) \quad (1)$$

where

$x(t)$  is an  $n$ -vector of state variables

$u(t)$  is an  $m$ -vector of control variables

$t$  is the independent variable (hereafter called time)

$f$  is an  $n$ -vector of known functions of  $x, u, t$

$(\dot{\phantom{x}})$  is  $\frac{d}{dt}(\phantom{x})$

The system operates over a finite time interval. The initial time  $t_0$  is assumed fixed, but the initial state is a vector of random variables with specified ensemble average properties. The problem is to minimize the ensemble average of a given function of the terminal conditions\*

$$J = \mathcal{E} \{ \varphi[x(t_f), t_f] \} \quad (2)$$

subject to the constraints

$$\mathcal{E} \{ \psi[x(t_f), t_f] \} = 0 \quad (3)$$

$$\mathcal{E} \{ (\psi^j[x(t_f), t_f])^2 \} = N^j \quad (4)$$

where  $g^j$  is the  $j^{\text{th}}$  component of any vector  $g$ .  $J$  is to be minimized while specifying the means and variances of the functions  $\psi^j$ .

---

\*  $t_f$  is the terminal time



It is assumed that nominal\* control programs  $\bar{u}(t)$  have been determined which minimize  $\varphi[\bar{x}(\bar{t}_f), \bar{t}_f]$  while meeting constraints  $\psi[\bar{x}(\bar{t}_f), \bar{t}_f] = 0$ . Thus,

$$\dot{\bar{x}} = f(\bar{x}, \bar{u}, t) \quad (5)$$

$$\dot{\lambda} = - \left( \frac{\partial f}{\partial x} \right)^T \lambda \quad (6)$$

$$\lambda^T \frac{\partial f}{\partial u} = 0 \quad \left( \frac{\partial^2 f}{\partial u^2} \neq 0 \text{ is assumed} \right) \quad (7)$$

with boundary conditions

$$t_0, \bar{x}(t_0) \text{ specified} \quad (8)$$

$$\psi[\bar{x}(\bar{t}_f), \bar{t}_f] = 0 \quad (9)$$

$$\lambda^T(\bar{t}_f) = \left( \frac{\partial \varphi}{\partial x} + \bar{\nu}^T \frac{\partial \psi}{\partial x} \right)_{t=\bar{t}_f} \quad (10)$$

$$(\lambda^T f)_{t=\bar{t}_f} = - \left( \frac{\partial \varphi}{\partial t} + \bar{\nu}^T \frac{\partial \psi}{\partial t} \right)_{t=\bar{t}_f} \quad (11)$$

where  $( )^T$  is the transpose of  $( )$ , the  $ij^{\text{th}}$  element of a matrix  $\frac{\partial g}{\partial y}$ ,  $g$  and  $y$  both vectors, is  $\frac{\partial g^i}{\partial y^j}$ . With  $\bar{u}(t)$  and  $\bar{x}(t)$  specified, the analysis will be carried out in terms of the perturbation quantities  $\delta u(t)$  and  $\delta x(t)$ , where, by definition

$$u(t) = \bar{u}(t) + \delta u(t) \quad (12)$$

$$x(t) = \bar{x}(t) + \delta x(t) \quad (13)$$

---

\*  $(\bar{\phantom{x}})$  is  $(\phantom{x})$  evaluated on the nominal path.

The minimization of  $J$  is to be carried out with respect to a number of parameters of the problem. One set of these parameters appears in the rule for terminating trajectories which must be imposed because there is no automatic way to determine  $t_f$  on each member of the ensemble. Suppose that the termination rule is described by

$$\Omega[x(t), t]_{t=t_f} = 0 \quad (14)$$

where  $\Omega$  may be any once differentiable function of  $x$  and  $t$ . Consistent with the second order approximation theory to be employed, the optimality of the reference trajectory leads to the result that the most general  $\Omega$  relevant in the analysis is a linear function of  $x$  and  $t$ . To first order, then, (14) may be written as

$$0 = \Omega[\bar{x}(\bar{t}_f), \bar{t}_f] + \left[ \frac{\partial \Omega}{\partial x} \delta x \right]_{t=\bar{t}_f} + \dot{\Omega} dt_f \quad (15)$$

where, by definition,

$$\dot{\Omega} = \left( \frac{\partial \Omega}{\partial x} \dot{x} + \frac{\partial \Omega}{\partial t} \right)_{t=\bar{t}_f}$$

Since  $t_f = \bar{t}_f + dt_f$ , the terminal time may be determined from (15) provided  $\dot{\Omega} \neq 0$ . This is simply the statement that  $\Omega$  must not be a constant of the motion if (15) is to give a solution for  $t_f$ .

Solving (15) for  $dt_f$  gives

$$dt_f = \bar{\Omega} + \Omega_x \delta x(\bar{t}_f) \quad (16)$$

where, by definition,

$$\bar{\Omega} = \frac{\Omega[\bar{x}(\bar{t}_f), \bar{t}_f]}{(-\dot{\Omega})}, \quad \Omega_x = \frac{\left( \frac{\partial \Omega}{\partial x} \right)_{t=\bar{t}_f}}{(-\dot{\Omega})}$$

$\bar{\Omega}$  and  $\Omega_x$  are the parameters to be optimized; it is evident there is no loss of generality in assuming  $\dot{\Omega} = -1$ .

The system controls for each member of the ensemble are assumed to satisfy

$$u(t) = \sum_{i=0}^{N_u} a_i t^i + \sum_{j=0}^{N_g} b_j t^j \left[ x(t) - \sum_{k=0}^{N_x} c_k t^k \right] \quad (17)$$

where  $N_u$ ,  $N_g$  and  $N_x$  are specified,  $a_i$ ,  $b_j$ ,  $c_k$  are unspecified. The first term in (17) is the polynomial approximation to  $\bar{u}(t)$ . The second term is the result of an assumption that the feedback control is linear in  $x(t)$ . The  $\sum c_k t^k$  is the polynomial approximation to  $\bar{x}(t)$ .  $\sum b_j t^j$  is the assumed form of the feedback gain. The most general linear feedback would use an  $m \times n$  matrix, say  $\Lambda(t)$ , of unspecified functions of time. Thus, the formulation used here replaces the most general linear feedback control system, which would require storage of  $\bar{u}(t)$ ,  $\bar{x}(t)$  and  $\Lambda(t)$ , by a linear feedback control utilizing polynomial approximations. It may be verified by inspection that  $a_i$ ,  $b_j$ ,  $c_k$  are  $m \times 1$ ,  $m \times n$ ,  $n \times 1$  matrices for each  $i$ ,  $j$ ,  $k$  respectively.

The problem, then, is to simultaneously choose all parameters  $\bar{\Omega}$ ,  $\Omega_x$ ,  $a$ ,  $b$ ,  $c$  to minimize  $J$  while satisfying the  $\psi^j$  mean and variance constraints.

## Derivation of Necessary Conditions for the Optimal Parameters

The approach used here will be to adjoin all relevant constraints to the performance index by means of Lagrange multipliers. Hence,

$$J = \mathcal{E}\{\varphi[x(t_f), t_f]\} + \nu^T \mathcal{E}\{\psi[x(t_f), t_f]\} + \frac{1}{2} \sum_j k_j \left[ \mathcal{E}\left\{\left(\psi^j[x(t_f), t_f]\right)^2\right\} - N^j \right] \\ + \mathcal{E} \int_{t_0}^{\bar{t}_f} (\lambda^T + \delta \lambda^T)(f - \dot{x}) dt \quad (18)$$

The essential approximation of the analysis is the assumption of "small" perturbations. The ensemble of system trajectories is treated by expanding about the nominal path and keeping terms through quadratic in  $\delta x$  and  $\delta u$ , but dropping higher order terms. As an example:\*

$$\mathcal{E}\{\varphi[x(t_f), t_f]\} = \varphi[\bar{x}(\bar{t}_f), \bar{t}_f] + \left[ \frac{\partial \varphi}{\partial x} \mathcal{E}(dx) + \frac{\partial \varphi}{\partial t} \mathcal{E}(dt) \right]_{t=\bar{t}_f} \\ + \frac{1}{2} \mathcal{E} \left[ dx^T \frac{\partial^2 \varphi}{\partial x^2} dx + dx^T \frac{\partial^2 \varphi}{\partial x \partial t} dt + dt \frac{\partial^2 \varphi}{\partial t \partial x} dx + dt \frac{\partial^2 \varphi}{\partial t^2} dt \right]_{t=\bar{t}_f} \quad (19)$$

Evaluation of (19) requires evaluation of

$$d[x(t_f)] = x(t_f) - \bar{x}(\bar{t}_f) \\ = \delta x(\bar{t}_f) + \int_{\bar{t}_f}^{t_f} \dot{x}(\tau) d\tau \quad (20)$$

But

---

\* The  $ij^{\text{th}}$  element of  $\partial^2 h / \partial y \partial z$ , where  $h$  is scalar and  $y$  and  $z$  are

vectors, is defined to be  $\frac{\partial^2 h}{\partial y^i \partial z^j}$ .

$$\begin{aligned}
\dot{x}(\tau) &= \dot{\bar{x}}(\tau) + \delta \dot{x}(\tau) \\
&= \dot{\bar{x}}(\bar{t}_f) + \ddot{\bar{x}}(\bar{t}_f)(\tau - \bar{t}_f) + \dots \\
&\quad + \left[ \frac{\partial f}{\partial x} \delta x + \frac{\partial f}{\partial u} \delta u \right]_{t=\bar{t}_f} + \dots
\end{aligned} \tag{21}$$

Substituting (21) into (20) and dropping terms above second order gives

$$\begin{aligned}
d[x(t_f)] &= \delta x(\bar{t}_f) + \dot{\bar{x}}(\bar{t}_f) dt_f + \frac{1}{2} \ddot{\bar{x}}(\bar{t}_f) dt_f^2 \\
&\quad + \left[ \frac{\partial f}{\partial x} \delta x + \frac{\partial f}{\partial u} \delta u \right]_{t=\bar{t}_f} dt_f
\end{aligned} \tag{22}$$

Everywhere in (19) that  $dt_f$  appears it is replaced by  $\bar{\Omega} + \Omega_x \delta x(\bar{t}_f)$ .<sup>\*</sup> This makes all terminal functions depend only on quantities evaluated at  $t = \bar{t}_f$ .

The  $\dot{x}$  terms in (18) are integrated by parts. The Lagrange multipliers  $\nu$  are written

$$\nu = \bar{\nu} + d\nu \tag{23}$$

where  $d\nu$  is assumed to be of order  $\delta x(\bar{t}_f)$ . It is further assumed that the Lagrange multiplier functions  $\delta \lambda(t)$  are of order  $\delta x(t)$ .<sup>\*\*</sup> The Lagrange multipliers  $k_j$  are assumed to be order one. These assumptions all rely on the basic assumption that the entire ensemble of trajectories lies within an adequately small neighborhood of the reference path.

Expansion of (18) through second order and grouping similar terms gives

---

\*  $\bar{\Omega}$  is assumed to be the order of  $\mathcal{E}[\delta x(\bar{t}_f)]$ .

\*\* Note that  $\delta \lambda(t)$  is different on each member of the ensemble, just as  $\delta x(t)$  is.  $\lambda(t)$  is the same for each member, given by (6) and (10).

$$\begin{aligned}
J = & \bar{\Phi} + \left[ \frac{\partial \Phi}{\partial x} (I + \dot{x} \Omega_x) + \frac{\partial \Phi}{\partial t} \Omega_x \right]_{t=\bar{t}_f} \mathcal{E}[\delta x(\bar{t}_f)] + \left[ \frac{\partial \Phi}{\partial x} \dot{x} + \frac{\partial \Phi}{\partial t} \right]_{t=\bar{t}_f} \bar{\Omega} \\
& + \mathcal{E} \left[ \frac{\partial \Phi}{\partial x} \frac{\partial f}{\partial u} \delta u (\bar{\Omega} + \Omega_x \delta x) \right]_{t=\bar{t}_f} + \frac{1}{2} \mathcal{E} \left\{ \delta x^T \left[ \frac{\partial^2 \Phi}{\partial x^2} + 2 \Omega_x^T \frac{\partial \Phi}{\partial x} \frac{\partial f}{\partial x} \right. \right. \\
& + \left( \frac{\partial \Phi}{\partial x} \right)^T \Omega_x + \Omega_x^T \left( \frac{\partial \Phi}{\partial x} \right) + \Omega_x^T \ddot{\Phi} \Omega_x + \sum_{j=1}^p k_j \left( \frac{\partial \psi^j}{\partial x} + \dot{\psi}^j \Omega_x \right) \left( \frac{\partial \psi^j}{\partial x} + \dot{\psi}^j \Omega_x \right) \Big] \\
& \delta x \Big\}_{t=\bar{t}_f} + \bar{\Omega} \left\{ \ddot{\Phi} \Omega_x + \frac{\partial \Phi}{\partial x} \frac{\partial f}{\partial x} + \left( \frac{\partial \Phi}{\partial x} \right)^T + \sum_{j=1}^p k_j \dot{\psi}^j \left( \frac{\partial \psi^j}{\partial x} + \dot{\psi}^j \Omega_x \right) \right\}_{t=\bar{t}_f} \\
& \mathcal{E}[\delta x(\bar{t}_f)] - \sum_{j=1}^p k_j N^j + \frac{1}{2} \bar{\Omega}^2 \left[ \ddot{\Phi} + \sum_{j=1}^p k_j (\dot{\psi}^j)^2 \right]_{t=\bar{t}_f} + d\nu^T \left[ \left( \frac{\partial \psi}{\partial x} + \dot{\psi} \Omega_x \right) \right. \\
& \left. \mathcal{E}(\delta x) + \dot{\psi} \bar{\Omega} \right]_{t=\bar{t}_f} - [\lambda^T \mathcal{E}(\delta x)]_{t=\bar{t}_f} + [\lambda^T \mathcal{E}(\delta x)]_{t=t_0} - \mathcal{E}[\delta \lambda^T \delta x]_{t=\bar{t}_f} \\
& + \mathcal{E}[\delta \lambda^T \delta x]_{t=t_0} + \mathcal{E} \int_{t_0}^{\bar{t}_f} \{ (\dot{\lambda}^T + \delta \dot{\lambda}^T)(\bar{x} + \delta x) + (\lambda^T + \delta \lambda^T) \bar{f} + \lambda^T \left( \frac{\partial f}{\partial x} \delta x \right. \right. \\
& + \frac{\partial f}{\partial u} \delta u) + \frac{1}{2} [\delta x^T \frac{\partial^2 H}{\partial x^2} \delta x + \delta x^T \frac{\partial^2 H}{\partial x \partial u} \delta u + \delta u^T \frac{\partial^2 H}{\partial u \partial x} \delta x \\
& + \delta u^T \frac{\partial^2 H}{\partial u^2} \delta u] + \delta \lambda^T \left( \frac{\partial f}{\partial x} \delta x + \frac{\partial f}{\partial u} \delta u \right) \} dt
\end{aligned} \tag{24}$$

where extensive use has been made of the following notational substitutions:

$$\Phi = \varphi + \bar{\nu}^T \psi$$

$$H = \lambda^T f$$

$I$  is the identity matrix

$$\dot{() } = \frac{\partial ( )}{\partial x} \dot{x} + \frac{\partial ( )}{\partial t}$$

$$\ddot{() } = \dot{x}^T \frac{\partial^2 ( )}{\partial x^2} \dot{x} + \dot{x}^T \frac{\partial^2 ( )}{\partial x \partial t} + \frac{\partial ( )}{\partial x} \ddot{x} + \frac{\partial^2 ( )}{\partial t \partial x} \dot{x} + \frac{\partial^2 ( )}{\partial t^2}$$

and all derivatives are evaluated on the reference path.

Using (12), (13), (17),  $\delta u(t)$  may be written as

$$\begin{aligned} \delta u(t) &= u(t) - \bar{u}(t) \\ &= \sum_{i=0}^{N_u} a_i t^i - \bar{u}(t) + \sum_{j=0}^{N_g} b_j t^j \left[ \delta x(t) + \bar{x}(t) - \sum_{k=0}^{N_x} c_k t^k \right] \end{aligned} \quad (25)$$

The following purely symbolic notations are introduced for convenience:

$$\sum_{i=0}^{N_u} a_i t^i = at$$

$$\sum_{j=0}^{N_g} b_j t^j = bt$$

$$\sum_{k=0}^{N_x} c_k t^k = ct$$

With these substitutions  $\delta u(t)$  may be written as

$$\delta u(t) = at - \bar{u}(t) + bt [\delta x(t) + \bar{x}(t) - ct] \quad (26)$$

$\delta u(t)$  from (26) may be substituted into (24), giving  $J$  as a function of  $\bar{\Omega}$ ,  $\Omega_x$ ,  $a$ ,  $b$ ,  $c$  and other quantities. A necessary condition for optimal choice of the unspecified parameters is that  $dJ$  be zero for arbitrary first order changes in the parameters.

By virtue of the optimality of the reference trajectory, all first order terms in  $dJ$ , and the  $\delta u(\bar{t}_f)$  term also, drop out. Thus,  $dJ$  is composed entirely of second order terms and by a straightforward development, may be written as

$$\begin{aligned}
dJ = & e \left\{ \delta x^T \left[ \frac{\partial^2 \Phi}{\partial x^2} + 2 \Omega_x^T \frac{\partial \Phi}{\partial x} \frac{\partial f}{\partial x} + \left( \frac{\partial \Phi}{\partial x} \right)^T \Omega_x + \Omega_x^T \left( \frac{\partial \Phi}{\partial x} \right) + \Omega_x^T \ddot{\Phi} \Omega_x \right. \right. \\
& + \sum_{j=1}^p k_j \left( \frac{\partial \psi^j}{\partial x} + \dot{\psi}^j \Omega_x \right)^T \left( \frac{\partial \psi^j}{\partial x} + \dot{\psi}^j \Omega_x \right) \left. + \bar{\Omega} \left[ \ddot{\Phi} \Omega_x + \frac{\partial \Phi}{\partial x} \frac{\partial f}{\partial x} + \left( \frac{\partial \Phi}{\partial x} \right)^T \right. \right. \\
& + \sum_{j=1}^p k_j \dot{\psi}^j \left( \frac{\partial \psi^j}{\partial x} + \dot{\psi}^j \Omega_x \right) \left. + d\nu^T \left[ \frac{\partial \psi}{\partial x} + \dot{\psi} \Omega_x \right] - \delta \lambda^T \right\} \delta(\delta x) \Big|_{t=\bar{t}_f} \\
& + \left\{ \left[ \ddot{\Phi} \Omega_x + \frac{\partial \Phi}{\partial x} \frac{\partial f}{\partial x} + \left( \frac{\partial \Phi}{\partial x} \right)^T + \sum_{j=1}^p k_j \dot{\psi}^j \left( \frac{\partial \psi^j}{\partial x} + \dot{\psi}^j \Omega_x \right) \right] e(\delta x) + \bar{\Omega} \left[ \ddot{\Phi} \right. \right. \\
& + \sum_{j=1}^p k_j (\dot{\psi}^j)^2 \left. \right] + d\nu^T \dot{\psi} \Big\}_{t=\bar{t}_f} d\bar{\Omega} + \text{tr} \left\{ X \left[ \left( \frac{\partial \Phi}{\partial x} \right)^T + \left( \frac{\partial f}{\partial x} \right)^T \left( \frac{\partial \Phi}{\partial x} \right)^T + \Omega_x^T \ddot{\Phi} \right. \right. \\
& + \sum_{j=1}^p k_j \left( \frac{\partial \psi^j}{\partial x} + \dot{\psi}^j \Omega_x \right)^T \dot{\psi}^j \left. + e(\delta x) \left[ \bar{\Omega} \ddot{\Phi} + \bar{\Omega} \sum_{j=1}^p k_j (\dot{\psi}^j)^2 + d\nu^T \dot{\psi} \right] \right\}_{t=\bar{t}_f} d\Omega_x \\
& + e \int_{t_0}^{\bar{t}_f} \left\{ \left[ \delta \dot{\lambda}^T + \delta \lambda^T \left( \frac{\partial f}{\partial x} + \frac{\partial f}{\partial u} b t \right) + \delta x^T \left( \frac{\partial^2 H}{\partial x^2} + \frac{\partial^2 H}{\partial x \partial u} b t + (b t)^T \frac{\partial^2 H}{\partial u \partial x} \right. \right. \right. \\
& + (b t)^T \frac{\partial^2 H}{\partial u^2} b t \left. \right] + [a t - \bar{u} + b t(\bar{x} - c t)]^T \left( \frac{\partial^2 H}{\partial u \partial x} + \frac{\partial^2 H}{\partial u^2} b t \right) \left. \right] \delta(\delta x) \\
& + \sum_{i=0}^{N_u} \left[ \delta x^T \frac{\partial^2 H}{\partial x \partial u} + [a t - \bar{u} + b t(\bar{x} - c t)]^T \frac{\partial^2 H}{\partial u^2} + (\delta x^T L + \epsilon^T) \frac{\partial f}{\partial u} \right] t^i da_i
\end{aligned}$$



$$\begin{aligned}
& + \sum_{j=0}^{N_g} \text{tr} \left[ (\delta x + \bar{x} - ct) \left( \delta x^T \frac{\partial^2 H}{\partial x \partial u} + [at - \bar{u} + bt(\delta x + \bar{x} - ct)]^T \frac{\partial^2 H}{\partial u^2} \right. \right. \\
& + \left. \left. (\delta x^T L + \iota^T) \frac{\partial f}{\partial u} \right) \right] t^j db_j - \sum_{k=0}^{N_x} \left[ \delta x^T \frac{\partial^2 H}{\partial x \partial u} + [at - \bar{u} + bt(\delta x + \bar{x} - ct)]^T \frac{\partial^2 H}{\partial u^2} \right. \\
& + \left. \left. (\delta x^T L + \iota^T) \frac{\partial f}{\partial u} \right] t^k dc_k \right\} dt
\end{aligned} \tag{27}$$

where, by definition,  $\text{tr}$  stands for trace and

$$X(t) = \mathcal{E}[\delta x(t) \delta x^T(t)]$$

Setting  $dJ = 0$  provides necessary conditions for extremizing choices of the control parameters. The Lagrange multipliers  $\delta \lambda$  satisfy

$$\begin{aligned}
& \delta \dot{\lambda} + \left( \frac{\partial f}{\partial x} + \frac{\partial f}{\partial u} bt \right)^T \delta \lambda + \left( \frac{\partial^2 H}{\partial x^2} + (bt)^T \frac{\partial^2 H}{\partial u \partial x} + \frac{\partial^2 H}{\partial x \partial u} bt + (bt)^T \frac{\partial^2 H}{\partial u^2} bt \right) \delta x \\
& + \left( \frac{\partial^2 H}{\partial x \partial u} + (bt)^T \frac{\partial^2 H}{\partial u^2} \right) [at - \bar{u} + bt(\bar{x} - ct)] = 0
\end{aligned} \tag{28}$$

$$\begin{aligned}
\delta \lambda(\bar{t}_f) = & \left\{ \left[ \frac{\partial^2 \Phi}{\partial x^2} + 2 \left( \frac{\partial f}{\partial x} \right)^T \left( \frac{\partial \Phi}{\partial x} \right)^T \Omega_x + \left( \frac{\partial \Phi}{\partial x} \right)^T \Omega_x + \Omega_x^T \left( \frac{\partial \Phi}{\partial x} \right) + \Omega_x^T \ddot{\Phi} \Omega_x \right. \right. \\
& + \sum_{j=1}^p k_j \left( \frac{\partial \psi^j}{\partial x} + \dot{\psi}^j \Omega_x \right)^T \left( \frac{\partial \psi^j}{\partial x} + \dot{\psi}^j \Omega_x \right) \Big] \delta x + \bar{\Omega} \left[ \ddot{\Phi} \Omega_x^T + \right. \\
& \left. \left. \left( \frac{\partial f}{\partial x} \right)^T \left( \frac{\partial \Phi}{\partial x} \right)^T + \left( \frac{\partial \Phi}{\partial x} \right)^T + \sum_{j=1}^p k_j \left( \frac{\partial \psi^j}{\partial x} + \dot{\psi}^j \Omega_x \right)^T \dot{\psi}^j \right] + \left( \frac{\partial \psi}{\partial x} + \dot{\psi} \Omega_x \right)^T d\nu \Big\}_{t=\bar{t}_f}
\end{aligned} \tag{29}$$

Because  $\delta x(t)$  is a vector of random variables,  $\delta \lambda(t)$  is also. Neither can be used computationally. However, it may be verified by direct substitution that

$$\delta \lambda(t) = L(t) \delta x(t) + \ell(t) \quad (30)$$

where  $L(t)$  and  $\ell(t)$  are the same for every member of the ensemble.

$$\begin{aligned} \dot{L} + \left( \frac{\partial f}{\partial x} + \frac{\partial f}{\partial u} bt \right)^T L + L \left( \frac{\partial f}{\partial x} + \frac{\partial f}{\partial u} bt \right) + \frac{\partial^2 H}{\partial x^2} + (bt)^T \frac{\partial^2 H}{\partial u \partial x} \\ + \frac{\partial^2 H}{\partial x \partial u} bt + (bt)^T \frac{\partial^2 H}{\partial u^2} bt = 0 \end{aligned} \quad (31)$$

$$\dot{\ell} + \left( \frac{\partial f}{\partial x} + \frac{\partial f}{\partial u} bt \right)^T \ell + \left( \frac{\partial^2 H}{\partial x \partial u} + (bt)^T \frac{\partial^2 H}{\partial u^2} \right) (at - \bar{u} + bt(\bar{x} - ct)) = 0 \quad (32)$$

The boundary conditions for  $L$  and  $\ell$  are evident by inspection of (29).

To obtain the remainder of the necessary conditions resulting from  $dJ = 0$ , it is necessary to develop the differential equations for  $\mathcal{E}(\delta x)$  and for  $X$ . First, it may be noted that  $\mathcal{E}(\delta x)$  appears only in terms that are second order, hence it need be calculated only to first order. The linearized perturbations of (1) with  $\delta u(t)$  from (26) immediately give

$$\frac{d}{dt} \mathcal{E}(\delta x) = \left( \frac{\partial f}{\partial x} + \frac{\partial f}{\partial u} bt \right) \mathcal{E}(\delta x) + \frac{\partial f}{\partial u} [at - \bar{u} + bt(\bar{x} - ct)] \quad (33)$$

It is convenient to define

$$\delta x = \mathcal{E}(\delta x) + \tilde{\delta x} \quad (34)$$

so that

$$X = \mathcal{E}(\delta x) \mathcal{E}(\delta x^T) + \tilde{X} \quad (35)$$

$$\tilde{X} = \mathcal{E}[\delta \tilde{x} \delta \tilde{x}^T] \quad (36)$$

Then, by direct substitution

$$\dot{\tilde{X}} = \left( \frac{\partial f}{\partial x} + \frac{\partial f}{\partial u} b t \right) \tilde{X} + \tilde{X} \left( \frac{\partial f}{\partial x} + \frac{\partial f}{\partial u} b t \right)^T \quad (37)$$

The boundary conditions for  $\mathcal{E}(\delta x)$  and  $\tilde{X}$  are given by

$$\mathcal{E}[\delta x(t_0)] , \text{ specified} \quad (38)$$

$$\mathcal{E}[\delta x(t_0) \delta x^T(t_0)] = X(t_0), \text{ specified} \quad (39)$$

There are thus  $2(n^2 + n)$  differential equations for  $\mathcal{E}(\delta x)$ ,  $\tilde{X}$ ,  $\ell$ ,  $L$  and corresponding boundary conditions, half at  $t_0$  and half at  $\bar{t}_f$ . The conditions at  $\bar{t}_f$  involve the Lagrange multipliers  $d\nu$  and  $k_j$ ; constraint equations (3) and (4) furnish the additional required  $2p$  relations.

From (16) it is clear that  $\bar{\Omega}$  is a bias in the choice of  $dt_f$ . Such a bias gives added flexibility because the differences of  $at$  and  $ct$  from  $\bar{u}$  and  $\bar{x}$  respectively cause  $\mathcal{E}[\delta x(t)]$  to be non-zero. Applying  $dJ = 0$ ,  $\bar{\Omega}$  may be explicitly solved for in terms of other parameters of the problem:

$$\bar{\Omega} = - \left\{ \frac{\left[ \ddot{\Phi} \Omega_x + \frac{\partial \Phi}{\partial x} \frac{\partial f}{\partial x} + \left( \frac{\partial \dot{\Phi}}{\partial x} \right) + \sum_{j=1}^p k_j \dot{\psi}^j \left( \frac{\partial \psi^j}{\partial x} + \dot{\psi}^j \Omega_x \right) \right] \mathcal{E}(\delta x) + d\nu^T \dot{\psi}}{\ddot{\Phi} + \sum_{j=1}^p k_j (\dot{\psi}^j)^2} \right\}_{t=\bar{t}_f} \quad (40)$$

Thus,  $\bar{\Omega}$  need not appear as an unknown in any numerical optimization procedure.

The parameters  $\Omega_x$  may also be solved for, from  $dJ = 0$ , in terms of quantities evaluated at  $t = \bar{t}_f$ :

$$\Omega_x = - \left\{ \frac{\left[ \bar{\Omega} \ddot{\Phi} + \bar{\Omega} \sum_{j=1}^p k_j (\dot{\psi}^j)^2 + d\nu^T \dot{\psi} \right] \mathcal{E}(\delta x^T) X^{-1} + \left( \frac{\partial \Phi}{\partial x} \right) + \frac{\partial \Phi}{\partial x} \frac{\partial f}{\partial x} + \sum_{j=1}^p k_j \dot{\psi}^j \frac{\partial \psi^j}{\partial x}}{\ddot{\Phi} + \sum_{j=1}^p k_j (\dot{\psi}^j)^2} \right\}_{t=\bar{t}_f} \quad (41)$$

After utilizing (40) and (41), the unspecified parameters are  $a, b, c$ . These must satisfy integral relations which result from  $dJ = 0$ , for arbitrary small changes  $da, db, dc$ .

$$0 = \int_{t_0}^{\bar{t}_f} \left\{ \mathcal{E}(\delta x^T) \frac{\partial^2 H}{\partial x \partial u} + [at - \bar{u} + bt(\bar{x} - ct)]^T \frac{\partial^2 H}{\partial u^2} + [\mathcal{E}(\delta x^T)L + \ell^T] \frac{\partial f}{\partial u} \right\} t^i dt \quad (42)$$

$i = 0, 1, 2, \dots, N_u$

$$0 = \int_{t_0}^{\bar{t}_f} \left\{ [X + (\bar{x} - ct) \mathcal{E}(\delta x^T)] \frac{\partial^2 H}{\partial x \partial u} + [X(bt)^T + \mathcal{E}(\delta x)[at - \bar{u} + bt(\bar{x} - ct)]^T \frac{\partial^2 H}{\partial u^2} \right. \\ \left. + (\bar{x} - ct) \mathcal{E}(\delta x^T)(bt)^T + (\bar{x} - ct)[at - \bar{u} + bt(\bar{x} - ct)]^T \frac{\partial^2 H}{\partial u^2} \right. \\ \left. + ([X + (\bar{x} - ct) \mathcal{E}(\delta x^T)]L + [\mathcal{E}(\delta x) + (\bar{x} - ct)]\ell^T) \frac{\partial f}{\partial u} \right\} t^j dt \quad (43)$$

$j = 0, 1, 2, \dots, N_g$

$$0 = \int_{t_0}^{\bar{t}_f} \left\{ \mathcal{E}(\delta x^T) \frac{\partial^2 H}{\partial x \partial u} + [at - \bar{u} + bt(\mathcal{E}(\delta x) + \bar{x} - ct)]^T \frac{\partial^2 H}{\partial u^2} + [\mathcal{E}(\delta x^T)L + \ell^T] \frac{\partial f}{\partial u} \right\} t^k dt \quad (44)$$

$$k = 0, 1, 2, \dots, N_x$$

The parameters  $a, b, c$  may not be eliminated algebraically because other quantities depend on them. The necessary conditions involving  $\mathcal{E}(\delta x), \tilde{X}, t, L, a, b, c$  are all interlocked. This is characteristic of dynamic system optimization problems with control parameters. Although such problems are seldom easy, the one considered here presents no new conceptual difficulties.

## An Alternative Approach to the Necessary Conditions Derivation

This analysis is based on second order expansions and is closely related to the second variation guidance schemes of Refs. 1 and 2. There, for a single perturbed trajectory, the second variation of the performance index is minimized subject to satisfaction of the  $\dot{x} = f$  and  $\psi = 0$  constraints. One proceeds by making stationary the function  $\Phi = \varphi + \bar{v}^T \psi$ , where properly chosen  $\bar{v}$  will lead to satisfaction of the terminal constraints. The second variation of  $\Phi$ , from Refs. 1 and 2, is

$$J_2 = \left[ dx^T \frac{\partial^2 \Phi}{\partial x^2} dx + dx^T \frac{\partial^2 \Phi}{\partial x \partial t} dt + dt \frac{\partial^2 \Phi}{\partial t \partial x} dx + dt \frac{\partial^2 \Phi}{\partial t^2} dt \right]_{t=\bar{t}_f} \\ + \int_{t_0}^{\bar{t}_f} \left[ \delta x^T \frac{\partial^2 H}{\partial x^2} \delta x + \delta x^T \frac{\partial^2 H}{\partial x \partial u} \delta u + \delta u^T \frac{\partial^2 H}{\partial u \partial x} \delta x + \delta u^T \frac{\partial^2 H}{\partial u^2} \delta u \right] dt \quad (45)$$

Since the reference path satisfies all the constraints, it is sufficient to adjoin the linearized perturbation constraints

$$\delta \dot{x} = \frac{\partial f}{\partial x} \delta x + \frac{\partial f}{\partial u} \delta u \quad (46)$$

$$d\psi = \left[ \frac{\partial \psi}{\partial x} dx + \frac{\partial \psi}{\partial t} dt \right]_{t=\bar{t}_f} = 0 \quad (47)$$

Then, given  $\delta x(t_0)$ ,  $\delta u(t)$  is chosen to minimize  $\frac{1}{2} J_2$  while satisfying constraints (46) and (47). This leads to a linear feedback relation

$$\delta u(t) = -\Lambda(t) \delta x(t) \quad (48)$$

It is tacitly assumed that  $\bar{x}(t)$  and  $\bar{u}(t)$  as well as  $\Lambda(t)$  are "stored" (available to the guidance system).

The significant operational simplification of neighboring extremal guidance introduced in this paper is the substitution of a relatively small number of polynomial coefficients for the functions  $\bar{u}(t)$ ,  $\bar{x}(t)$ ,  $\Lambda(t)$ . The general functions of time would require tables of values vs. time in operation with a digital computer. Use of polynomial coefficients instead may be expected to greatly reduce the storage requirements.

An additional advantage of the polynomial approximations is that the difficulty  $\Lambda(t) \rightarrow \infty$  as  $t \rightarrow \bar{t}_f$  disappears. The polynomial  $\sum b_j t^j$  will certainly be well behaved in the neighborhood of  $t = \bar{t}_f$ . Thus, the need for a transverse state comparison, so important for neighboring extremal control, may become less significant in analyses conducted along the present lines.

It is, of course, necessary to satisfy the constraint (46) in any (small perturbation) analysis. It is not possible, however, to satisfy (47) for arbitrary  $\delta x(t_0)$  with the polynomial approximations. Hence, the use of a statistical performance index is not only appropriate, but even unavoidable. The alternative approach to the derivation of the previous section is to consider minimizing the ensemble average of  $\frac{1}{2} J_2$ . Constraints on the mean and variance of the  $\psi^j$ 's [equations (3) and (4)] are imposed. Because these ensemble averages involve only the mean and covariance of  $\delta x(t)$ , it is sufficient to use the differential equations for  $\mathcal{E}(\delta x)$  and  $\tilde{X}$  in place of (46). Thus, (24) is fully equivalent to

$$\begin{aligned}
 J = & \mathcal{E} \left[ \frac{1}{2} J_2 \right] + d\nu^T \mathcal{E} \{ d\psi[x(t_f), t_f] \} + \sum_{j=1}^p \frac{1}{2} k_j \mathcal{E} \{ d\psi^j[x(t_f), t_f] \}^2 + \int_{t_0}^{\bar{t}_f} \left\{ \mathcal{E}^T \left[ \left( \frac{\partial f}{\partial x} + \frac{\partial f}{\partial u} b t \right) \mathcal{E}(\delta x) \right. \right. \\
 & \left. \left. + [at - \bar{u} + bt(\bar{x} - ct)] - \frac{d}{dt} \mathcal{E}(\delta x) \right] + \text{tr} \left[ \left( \frac{\partial f}{\partial x} + \frac{\partial f}{\partial u} b t \right) \tilde{X} + \tilde{X} \left( \frac{\partial f}{\partial x} + \frac{\partial f}{\partial u} b t \right)^T - \dot{\tilde{X}} \right] \right\} dt
 \end{aligned}
 \tag{49}$$

Here  $\lambda(t)$  and  $L(t)$  appear as a vector and matrix respectively of Lagrange multiplier functions.\*  $\lambda(t)$  is the vector adjoint to  $\delta\mathcal{E}[\delta x(t)]$ ,  $L(t)$  is the matrix adjoint to  $\delta X(t)$ . All the necessary conditions of the previous section may be obtained by requiring  $J$  of (49) to be stationary with respect to arbitrary small changes in the unspecified parameters.

---

\* Since  $L$  multiplies symmetric matrices in (49), it may be assumed symmetric with no loss of generality.



### Possible Additional Complexities

The analysis as presented allows disturbances only in the form of perturbations in the initial state variables. It also assumes that the state is known perfectly at all times. Both restrictions may be relaxed while still retaining the polynomial approximation approach.

Disturbing influences may arise from perturbations of system parameters from their reference values. For example, the thrust and/or fuel consumption rate of a rocket vehicle may deviate from its pre-planned value. To allow for this in the analysis presented here, such system parameters may be regarded as state variables with zero time derivatives. Thus, a parameter deviation becomes an initial state variable perturbation.

Time-dependent random forcing functions may be added to the analysis if their means and covariances are known, although serious complications may arise if the noise is appreciably correlated in time. The main effect with zero-mean white noise would be to add a term to  $\dot{X}$ . The other equations would be unaltered, but any numerical solution might be substantially different.

If state estimation errors were not considered negligible, it would be possible to include them by considering the estimator characteristics. A linear perturbation estimator would be consistent with the degree of approximation used here. The estimator gain matrix would play a role analogous to the feedback gain matrix. It would be approximated by a polynomial analogous to  $\sum_j b_j t^j$ . The polynomial coefficients would be added to the others, all to be chosen simultaneously to optimize the system ensemble average performance.

## Computational Considerations

The preceding analysis has been devoted to problem formulation and development of first order necessary conditions for a minimum. Computational determination of the control parameters which actually furnish a minimum represents a second phase of study. It is clear, however, that any of the methods applicable to the solution of Mayer/Bolza variational problems appear likely to be equally suitable to parameter optimization problems of the present type. On the basis of experience, the writers are favorably inclined toward the use of gradient methods (Refs. 7 and 8) and methods of the second variation type (Ref. 9), and in this connection it should be noted that the usual requirement for rapid access storage of control variables versus time is eased in favor of a somewhat less severe requirement for storage of parameter values. With the second order method of Ref. 9, it appears that parameter optimization will entail the solution of fairly large linear algebraic systems, and hence that greater attention than usual must be given to error propagation problems.

## Concluding Remarks

The present paper has sketched in some detail an ensemble averaging approach to optimal guidance polynomial approximations. Conclusions on the merits of this approach must be deferred until numerical examples of synthesis procedure have been worked and system simulations performed. In connection with the problem of guidance system mechanization, it will be of interest to investigate the use of transverse state comparison or some similar mode of comparison employing polynomial representation.

## References

1. Kelley, H.J.; "Guidance Theory and Extremal Fields," IRE National Aerospace Electronics Conference, May 14-16, 1962; also IRE Transactions on Automatic Control, October 1962.
2. Breakwell, J.V. and Bryson, A.E.; "Neighboring-Optimum Terminal Control for Multivariable Nonlinear Systems," SIAM Symposium on Multivariable System Theory, Cambridge, Mass., Nov. 1-3, 1962; also Breakwell, J.V., Speyer, J.L. and Bryson, A.E.; "Optimization and Control of Nonlinear Systems Using the Second Variation," JSIAM Control, Vol. 1, No. 2, 1963.
3. Kelley, H.J.; "An Optimal Guidance Approximation Theory," IEEE Transactions on Automatic Control, October 1964.
4. Dunn, J.C.; "Autonomous Time-Optimal Extremal Fields," to appear.
5. Striebel, C.T. and Breakwell, J.V.; "Minimum Effort Control in Interplanetary Guidance," IAS 31st Annual Meeting, New York, N.Y., January 21-23, 1963.
6. Denham, W.F. and Speyer, J.L.; "Optimal Measurement and Velocity Correction Programs for Midcourse Guidance," AIAA Summer Meeting, Los Angeles, Calif., June 17-20, 1963; also AIAA Journal, Vol. 2, No. 5, May 1964.
7. Kelley, H.J.; "Method of Gradients," Chapter 6 of Optimization Techniques, G. Leitmann, Editor, Academic Press, New York, 1962.
8. Bryson, A.E. and Denham, W.F.; "A Steepest-Ascent Method for Solving Optimum Programming Problems," Journal of Applied Mechanics, June 1962.
9. Kelley, H.J., Kopp, R.E. and Moyer, H.G.; "A Trajectory Optimization Technique Based Upon the Theory of the Second Variation," AIAA Astrodynamics Conference, New Haven, Conn., August 19-21, 1963. Also in Progress in Astronautics and Aeronautics, Vol. 14, Academic Press, New York, 1964.

N65 33054

NUMERICAL INVESTIGATION OF  
MINIMUM IMPULSE ORBITAL TRANSFER

Prepared by

Gary A. McCue and David F. Bender

Space Sciences Laboratory  
Space and Information Systems Division  
North American Aviation, Inc.

Special Report No. 7

October 1, 1964

Contract NAS8-5211

Prepared for

George C. Marshall Space Flight Center  
National Aeronautics and Space Administration  
Huntsville, Alabama

**NORTH AMERICAN AVIATION, INC.**  
**SPACE and INFORMATION SYSTEMS DIVISION**

## TECHNICAL REPORT INDEX/ABSTRACT

ACCESSION NUMBER	69496-64	DOCUMENT SECURITY CLASSIFICATION	Unclassified
TITLE OF DOCUMENT		LIBRARY USE ONLY	
Numerical Investigation of Minimum Impulse Orbital Transfer			
AUTHOR(S)			
Gary A. McCue and David F. Bender			
CODE	ORIGINATING AGENCY AND OTHER SOURCES	DOCUMENT NUMBER	
	NAA - S&ID	SID 64-423	
PUBLICATION DATE	CONTRACT NUMBER		
DESCRIPTIVE TERMS			

## ABSTRACT

A method for the precise numerical determination of optimum two-impulse orbital transfers between inclined elliptical orbits is described. A numerical optimization technique termed "adaptive steepest descent" is shown to overcome convergence difficulties encountered with other less powerful methods. A double-precision IBM 7094 program incorporating this technique was used to make detailed studies of complicated impulse function spaces associated with various orbit pairs which had been previously investigated by function contouring. Several interesting results, including the existence of at least three locally minimum two-impulse transfers between "almost tangent" coplanar elliptical orbits, were revealed by these numerical studies. Other numerical data indicates that minimum impulse transfer circumstances for coplanar orbits may be extended to strongly inclined orbits; that is, one and two-impulse maneuvers which are optimal exist for classes of inclined orbits. Results of a study of orbits which osculate to a Lawden Spiral are also presented. It was found that two-impulse transfers between these orbits always required less velocity change than a transfer between the points of osculation, along a Lawden Spiral. Extensive numerical comparison revealed that the difference in velocity change for the two maneuvers increases approximately as the 4.7 power of a parameter denoting distance between the points of osculation on the Lawden Spiral.

Author



## CONTENTS

	Page
I. INTRODUCTION . . . . .	1
II. TWO-IMPULSE ORBITAL TRANSFER FORMULATION . . . . .	2
Transfer Geometry . . . . .	2
Impulse Computation . . . . .	4
Impulse Minimization . . . . .	5
III. ADAPTIVE STEEPEST DESCENT . . . . .	10
Convergence Properties . . . . .	12
IV. NUMERICAL RESULTS . . . . .	17
"Almost Tangent" Orbits . . . . .	17
Inclined Elliptical Orbits . . . . .	20
Lawden Spiral vs. Two Impulse Transfer . . . . .	22
V. CONCLUSION . . . . .	24
VI. REFERENCES . . . . .	25



## ILLUSTRATIONS

Figure		Page
1	Descent Paths Plotted on Optimum-Impulse Contour Map	13
2	Optimization Process(Parameter Variation)	15
3	Coordinate Vector Behavior During Optimization Process	16
4	Optimum Impulse Contour Maps for Tangent Orbits	18
5	Optimum Impulse Variation Throughout Function Spaces-- "Almost Tangent" Orbits	19
6	Impulse Comparison for Optimum One and Two-Impulse Transfers	21
7	Lawden Spiral $\Delta V$ Compared to Optimum Two-Impulse $\Delta V$	23



## NOMENCLATURE

Scalars

$a$	Semimajor axis
$\alpha$	Magnitude of step size
$e$	Eccentricity (magnitude of $\underline{e}$ )
$i$	Inclination
$p$	Semilatus rectum
$r$	Radius to satellite
$s_1, s_2, s_3$	Scaling parameters
$\Delta\theta$	Transfer angle (true anomaly difference in transfer orbit plane)
$\mu$	Gravitation constant (95634.50100 mi <sup>3</sup> /sec <sup>2</sup> )
$\phi_1$	Angle from reference axis to departure position in initial orbit
$\phi_2$	Angle from reference axis to arrival position in terminal orbit
$\omega$	Argument of perigee, angle from reference axis to perigee point

Vectors

$\underline{e}$	Orbit shape and orientation vector
$\underline{g}$	Unit vector in gradient direction
$\underline{I}$	Impulse vector
$\underline{N}$	Unit vector denoting reference direction (line of intersection of initial and final orbit planes)
$\underline{r}$	Geocentric satellite position vector
$\underline{U}_1$	Unit vector directed toward point of departure from initial orbit



Vectors

$\underline{U}_2$	Unit vector directed toward point of arrival in final orbit
$\underline{V}$	Velocity vector
$\underline{W}$	Unit vector directed along orbit's angular momentum vector

Subscripts

1	Initial orbit
2	Final orbit
t	Transfer orbit
t1	transfer orbit departure point
t2	transfer orbit arrival point



## I. INTRODUCTION

In previous papers, (1, 2, 3) the authors discussed the properties of function spaces associated with optimum two-impulse transfer between inclined elliptical orbits. An impulse function contouring technique which presented the nature and structure of the entire function space was utilized to identify all possible regions of a given function which would yield optimum transfer orbits.

Contouring proves adequate for locating minima, and for providing insight, but it generally does not provide required numerical accuracy. This is true for many of the most interesting orbit pairs wherein the difficult phase of numerical optimization occurs during the final convergence. These particular functions are comprised of long, narrow "valleys" containing one or more minima. It is therefore necessary to employ an alternate technique to compute precise optimum orbital transfer circumstances for use in engineering design studies.

Experience with ordinary steepest descent processes (4) led to numerous frustrations and amplified the need for the more powerful adaptive steepest descent technique presented here. This rapid numerical method has been applied successfully to the minimization of numerous different orbital transfer function spaces. The method also has obvious application to a large class of problems which require numerical determination of the extrema of a function of 3 or more variables.



## II. TWO-IMPULSE ORBITAL TRANSFER FORMULATION

Adopting the notation of Ref. 1, consider a two-impulse transfer process between an initial orbit with elements  $p_1, e_1, \omega_1, i$  and a final orbit defined by  $p_2, e_2, \omega_2$ . The formulation assumes Keplerian orbits and results from choosing the final orbit as the reference plane;  $i$  is the relative inclination of the two orbit planes ( $\cos i = \underline{W}_1 \cdot \underline{W}_2$ , where  $\underline{W}_1$  and  $\underline{W}_2$  are unit vectors directed along the angular momentum vectors of the initial and final orbits). For coplanar orbits, the reference direction ( $\underline{N}$ ) is arbitrary; but for inclined orbits  $\underline{N}$  is defined as the line of intersection of the two orbit planes ( $\underline{N} = \underline{W}_2 \times \underline{W}_1 / |\underline{W}_2 \times \underline{W}_1|$ ).

For the general case, there is a three-parameter family of transfer orbits joining any two specific orbits. The angles from the reference line to departure point ( $\phi_1$ ) and to arrival point ( $\phi_2$ ) are a natural choice for two of the three independent variables, since they, along with the given orbital elements, specify position and velocity in the known orbits (Fig. 1). The semilatus rectum ( $p_t$ ) of the transfer orbit was the third parameter used for this study. It was chosen since it simplified the structure of the impulse function,  $I(\phi_1, \phi_2, p_t)$ . (5)

### TRANSFER GEOMETRY

Unit vectors ( $\underline{U}_1$  and  $\underline{U}_2$ ) and radius vectors ( $\underline{r}_1$  and  $\underline{r}_2$ ) toward the departure and arrival points may be computed from  $\phi_1, \phi_2$  and the elements of the initial and final orbits:\*

$$\underline{U}_1 = [\cos \phi_1, \sin \phi_1 \cos i, \sin \phi_1 \sin i] \quad (1)$$

---

\*The subscripts 1, 2, and t denote initial, final and transfer orbits.



$$\underline{U}_2 = [\cos \phi_2, \sin \phi_2, 0] \quad (2)$$

$$\underline{r}_j = \left[ \frac{p_j}{1 + e_j \cos (\phi_j - \omega_j)} \right] \underline{U}_j \quad j = 1, 2 \quad (3)$$

Unit vectors normal to the three orbit planes are defined as follows:

$$\underline{W}_1 = [0, -\sin i, \cos i] \quad (4)$$

$$\underline{W}_2 = [0, 0, 1] \quad (5)$$

$$\underline{W}_t = \underline{U}_1 \times \underline{U}_2 / |\underline{U}_1 \times \underline{U}_2| \quad \underline{U}_1 \times \underline{U}_2 \neq 0 \quad (6)$$

Two vectors that define the shape and orientation of the initial and final orbits complete the transfer geometry: (6)

$$\underline{e}_j = e_j [\cos \omega_j, \sin \omega_j \cos i_j, \sin \omega_j \sin i_j] \quad j = 1, 2 \quad (7)$$

The true anomaly interval traversed in the transfer orbit ( $\Delta \theta$ ) may be determined directly:

$$\cos \Delta \theta = (\underline{U}_1 \cdot \underline{U}_2) \quad 0^\circ < \Delta \theta < 180^\circ \quad (8)$$

No generality is lost if the true anomaly interval is limited to the first two quadrants. Although this does restrict the problem to "short transfers", if the signs of the velocity vectors in the transfer orbit are changed, the "long transfers" may be computed. Thus, in order to determine the absolute minimum impulse transfer between two elliptical orbits, it is necessary to compare the minima found from all the short transfers and all the long transfers.



## IMPULSE COMPUTATION

The function to be minimized is the total impulse for the two-impulse maneuver:

$$I = |\underline{I}_1| + |\underline{I}_2| \quad (9)$$

where

$$\underline{I}_1 = \pm \underline{V}_{t1} - \underline{V}_1 \quad (10)$$

$$\underline{I}_2 = \underline{V}_2 \mp \underline{V}_{t2} \quad (11)$$

(When a double sign is used, the upper sign refers to a "short transfer").

Velocity vectors in the initial and final orbits at the departure and arrival points ( $\underline{V}_1$  and  $\underline{V}_2$ ) and the corresponding velocity vectors in the transfer orbit ( $\underline{V}_{t1}$  and  $\underline{V}_{t2}$ ) are computed as follows:

$$\underline{V}_1 = \sqrt{\frac{\mu}{p_1}} \underline{W}_1 \times (\underline{e}_1 + \underline{U}_1) \quad (12)$$

$$\underline{V}_2 = \sqrt{\frac{\mu}{p_2}} \underline{W}_2 \times (\underline{e}_2 + \underline{U}_2) \quad (13)$$

$$\underline{V}_{t1} = \sqrt{\frac{\mu}{p_t}} (\underline{v} + z\underline{U}_1) \quad (14)$$

$$\underline{V}_{t2} = \sqrt{\frac{\mu}{p_t}} (\underline{v} - z\underline{U}_2) \quad (15)$$



where,

$$\underline{v} = \left[ \sqrt{\mu P_t} (\underline{r}_2 - \underline{r}_1) \right] / |\underline{r}_1 \times \underline{r}_2| \quad (16)$$

$$z = \frac{\sqrt{\mu}}{P_t} \tan (\Delta \theta / 2) \quad (17)$$

Eqs. 12 - 17 may be derived from Eq. 3.26 of Herget.<sup>(6)</sup> The final impulse equations are obtained from Eqs. 10 - 17 by substituting Eq. 6 and performing several algebraic manipulations:

$$\underline{I}_1 = \pm \left[ \underline{v} + z \underline{U}_1 \right] - \underline{V}_1 \quad (18)$$

$$\underline{I}_2 = \underline{V}_2 \mp \left[ \underline{v} - z \underline{U}_2 \right] \quad (19)$$

Impulses corresponding to long and short transfers are compared, and the combination producing the lesser impulse is used for the remaining computations. Because of the nature of the particular functions being analyzed, regions neighboring each local minimum are usually comprised entirely of either long or short transfers.

#### IMPULSE MINIMIZATION

Minimization of Eq. 9 by a steepest descent technique requires computation of the gradient. Upon differentiation, Eq. 9 provides the following expressions:

$$dI = \frac{(\underline{I}_1 \cdot d\underline{I}_1)}{|\underline{I}_1|} + \frac{(\underline{I}_2 \cdot d\underline{I}_2)}{|\underline{I}_2|} \quad (20)$$



or,

$$\frac{\partial I}{\partial p_t} = \frac{\left( I_1 \cdot \frac{\partial I_1}{\partial p_t} \right)}{|I_1|} + \frac{\left( I_2 \cdot \frac{\partial I_2}{\partial p_t} \right)}{|I_2|} \quad (21)$$

$$\frac{\partial I}{\partial \phi_1} = \frac{\left( I_1 \cdot \frac{\partial I_1}{\partial \phi_1} \right)}{|I_1|} + \frac{\left( I_2 \cdot \frac{\partial I_2}{\partial \phi_1} \right)}{|I_2|} \quad (22)$$

$$\frac{\partial I}{\partial \phi_2} = \frac{\left( I_1 \cdot \frac{\partial I_1}{\partial \phi_2} \right)}{|I_1|} + \frac{\left( I_2 \cdot \frac{\partial I_2}{\partial \phi_2} \right)}{|I_2|} \quad (23)$$

The above expressions may be expanded as follows:

$$\frac{\partial I_1}{\partial p_t} = \pm \frac{\partial V_{t1}}{\partial p_t} - \frac{\partial V_1}{\partial p_t} \quad (24)$$

$$\frac{\partial I_2}{\partial p_t} = \mp \frac{\partial V_{t2}}{\partial p_t} + \frac{\partial V_2}{\partial p_t} \quad (25)$$

$$\frac{\partial I_1}{\partial \phi_1} = \pm \frac{\partial V_{t1}}{\partial \phi_1} - \frac{\partial V_1}{\partial \phi_1} \quad (26)$$

$$\frac{\partial I_1}{\partial \phi_2} = \pm \frac{\partial V_{t1}}{\partial \phi_2} - \frac{\partial V_1}{\partial \phi_2} \quad (27)$$

$$\frac{\partial I_2}{\partial \phi_1} = \mp \frac{\partial V_{t2}}{\partial \phi_1} + \frac{\partial V_2}{\partial \phi_1} \quad (28)$$



$$\frac{\partial \underline{I}_2}{\partial \phi_2} = + \frac{\partial \underline{V}_{t2}}{\partial \phi_2} + \frac{\partial \underline{V}_2}{\partial \phi_2} \quad (29)$$

Noting that  $\partial \underline{V}_1 / \partial p_t$  and  $\partial \underline{V}_2 / \partial p_t$  are each zero, a simplified expression for  $\partial \underline{I} / \partial p_t$  may be obtained through several algebraic manipulations:

$$\frac{\partial \underline{I}}{\partial p_t} = + \frac{1}{2p_t} \left[ \frac{\underline{I}_1 \cdot (\underline{v} - z\underline{U}_1)}{|\underline{I}_1|} - \frac{\underline{I}_2 \cdot (\underline{v} + z\underline{U}_2)}{|\underline{I}_2|} \right] \quad (30)$$

Additional expressions are obtained from eqs. 26 - 29 by direct differentiation of the vector equations.

$$\begin{aligned} \frac{\partial \underline{V}_{t1}}{\partial \phi_1} = & \sqrt{\frac{\mu}{p_t}} \left\{ \frac{p_t}{r_1} \underline{U}_2 \left[ \frac{\partial \csc \Delta \theta}{\partial \phi_1} - \frac{\csc \Delta \theta}{r_1} \frac{\partial r_1}{\partial \phi_1} \right] \right. \\ & + \left( 1 - \frac{p_t}{r_2} \right) \left[ \underline{U}_1 \frac{\partial \csc \Delta \theta}{\partial \phi_1} + \csc \Delta \theta \frac{\partial \underline{U}_1}{\partial \phi_1} \right] \\ & \left. - \underline{U}_1 \frac{\partial \cot \Delta \theta}{\partial \phi_1} - \cot \Delta \theta \frac{\partial \underline{U}_1}{\partial \phi_1} \right\} \quad (31) \end{aligned}$$

$$\frac{\partial \underline{V}_1}{\partial \phi_1} = \sqrt{\frac{\mu}{p_1}} [-\cos \phi_1, -\sin \phi_1 \cos i, -\sin \phi_1 \sin i] \quad (32)$$

$$\begin{aligned} \frac{\partial \underline{V}_{t1}}{\partial \phi_2} = & \sqrt{\frac{\mu}{p_t}} \left\{ \frac{p_t}{r_1} \left[ \underline{U}_2 \frac{\partial \csc \Delta \theta}{\partial \phi_2} + \csc \Delta \theta \frac{\partial \underline{U}_2}{\partial \phi_2} \right] \right. \\ & + \underline{U}_1 \left[ \left( 1 - \frac{p_t}{r_2} \right) \frac{\partial \csc \Delta \theta}{\partial \phi_2} + \frac{p_t}{r_2^2} \csc \Delta \theta \frac{\partial r_2}{\partial \phi_2} \right. \\ & \left. \left. - \frac{\partial \cot \Delta \theta}{\partial \phi_2} \right] \right\} \quad (33) \end{aligned}$$

$$\frac{\partial \underline{V}_1}{\partial \phi_2} = 0 \quad (34)$$





$$\begin{aligned} \frac{\partial V_{t2}}{\partial \phi_1} = & \sqrt{\frac{\mu}{p_t}} \left\{ -\frac{p_t}{r_2} \left[ \underline{U}_1 \frac{\partial \csc \Delta \theta}{\partial \phi_1} + \csc \Delta \theta \frac{\partial \underline{U}_1}{\partial \phi_1} \right] \right. \\ & + \underline{U}_2 \left[ \left( 1 - \frac{p_t}{r_1} \right) \frac{\partial \csc \Delta \theta}{\partial \phi_1} - \frac{p_t}{r_1^2} \csc \Delta \theta \frac{\partial r_1}{\partial \phi_1} \right. \\ & \left. \left. + \frac{\partial \cot \Delta \theta}{\partial \phi_1} \right] \right\} \end{aligned} \quad (35)$$

$$\frac{\partial V_2}{\partial \phi_1} = 0 \quad (36)$$

$$\begin{aligned} \frac{\partial V_{t2}}{\partial \phi_2} = & \sqrt{\frac{\mu}{p_t}} \left\{ \frac{p_t}{r_2} \underline{U}_1 \left[ -\frac{\partial \csc \Delta \theta}{\partial \phi_2} + \frac{1}{r_2} \csc \Delta \theta \frac{\partial r_2}{\partial \phi_2} \right] \right. \\ & - \left( 1 - \frac{p_t}{r_1} \right) \left[ \underline{U}_2 \frac{\partial \csc \Delta \theta}{\partial \phi_2} + \csc \Delta \theta \frac{\partial \underline{U}_2}{\partial \phi_2} \right] \\ & \left. + \underline{U}_2 \frac{\partial \cot \Delta \theta}{\partial \phi_2} + \cot \Delta \theta \frac{\partial \underline{U}_2}{\partial \phi_2} \right\} \end{aligned} \quad (37)$$

$$\frac{\partial \underline{V}}{\partial \phi_2} = \sqrt{\frac{\mu}{p_2}} [-\cos \phi_2, -\sin \phi_2, 0] \quad (38)$$

The remaining undefined terms in Eqs. 31 - 38 may be computed from the following expressions:

$$\frac{\partial \underline{U}_1}{\partial \phi_1} = [-\sin \phi_1, \cos \phi_1 \cos i, \cos \phi_1 \sin i] \quad (39)$$

$$\frac{\partial \underline{U}_2}{\partial \phi_2} = [-\sin \phi_2, \cos \phi_2, 0] \quad (40)$$

$$\frac{\partial r_1}{\partial \phi_1} = \frac{r_1^2 e_1 \sin(\phi_1 - \omega_1)}{p_1} \quad (41)$$

$$\frac{\partial r_2}{\partial \phi_2} = \frac{r_2^2 e_2 \sin(\phi_2 - \omega_2)}{p_2} \quad (42)$$

$$\frac{\partial \csc \Delta \theta}{\partial \phi_j} = -\csc \Delta \theta \cot \Delta \theta \frac{\partial \Delta \theta}{\partial \phi_j} \quad j = 1, 2 \quad (43)$$



$$\frac{\partial \cot \Delta \theta}{\partial \phi_j} = - \csc^2 \Delta \theta \frac{\partial \Delta \theta}{\partial \phi_j} \quad j = 1, 2 \quad (44)$$

The following convenient expression for  $\Delta \theta$  allows computation of the remaining derivatives:

$$\Delta \theta = \cos^{-1} (\cos \phi_1 \cos \phi_2 + \sin \phi_1 \sin \phi_2 \cos i) \quad (45)$$

$$\frac{\partial \Delta \theta}{\partial \phi_1} = \frac{-(-\sin \phi_1 \cos \phi_2 + \cos \phi_1 \sin \phi_2 \cos i)}{\sqrt{1 - (\cos \phi_1 \cos \phi_2 + \sin \phi_1 \sin \phi_2 \cos i)^2}} \quad (46)$$

$$\frac{\partial \Delta \theta}{\partial \phi_2} = \frac{-(-\cos \phi_1 \sin \phi_2 + \sin \phi_1 \cos \phi_2 \cos i)}{\sqrt{1 - (\cos \phi_1 \cos \phi_2 + \sin \phi_1 \sin \phi_2 \cos i)^2}} \quad (47)$$



### III. ADAPTIVE STEEPEST DESCENT

Since setting Eqs. 21, 22, and 23 equal to zero yields no general analytical solution, one is faced with numerically solving an ordinary calculus problem requiring the minimization of a function of 3 variables.

Successful use of a numerical search which stepped in the negative gradient direction was reported in Refs. 4, 7, 8, and 9. However, this procedure proved to be inadequate for the more sensitive function spaces. Attempts to employ Newton-Raphson methods were similarly frustrated by the nature, structure, and multiplicity of minima of typical impulse function spaces.

The present "adaptive steepest descent" procedure effectively overcomes the convergence and accuracy limitations of the previous methods. A numerical search employing Eqs. 1 - 47 is terminated when the following necessary conditions for a local minimum have been achieved:

$$\frac{\partial I}{\partial \phi_1} \leq \epsilon, \quad \frac{\partial I}{\partial \phi_2} \leq \epsilon, \quad \frac{\partial I}{\partial P_t} \leq \epsilon, \quad \epsilon \ll 0 \quad (48)$$

During the  $n$ 'th step of the search the gradient vector is computed and the  $n + 1$ 'st coordinate vector is determined as follows:

$$\begin{bmatrix} \phi_1 \\ \phi_2 \\ P_t \end{bmatrix}_{n+1} = \begin{bmatrix} \phi_1 \\ \phi_2 \\ P_t \end{bmatrix}_n - \alpha \begin{bmatrix} \frac{s_1}{s_j} & 0 & 0 \\ 0 & \frac{s_2}{s_j} & 0 \\ 0 & 0 & \frac{s_3}{s_j} \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix} \quad j = 1, 2, \text{ or } 3 \quad (49)$$

where  $\alpha$  is the current magnitude of the step size, the  $s_j$  are variable scaling parameters, and  $g_1$ ,  $g_2$ , and  $g_3$  are the components of a unit vector



in the gradient direction. Note also that the scaling matrix is normalized relative to one of the scaling parameters. Eq. 49 is employed to construct a sequence of points at which the impulse function,  $I(\phi_1, \phi_2, p_t)$ , is evaluated. An additional constraint on the process requires that the sequence of impulses  $\{I(\phi_1, \phi_2, p_t)_n\}$  be monotone decreasing.

The control logic for the optimization process is rather simple:

- (1) If the inequality

$$I(\phi_1, \phi_2, p_t)_{n+1} < I(\phi_1, \phi_2, p_t)_n \quad (50)$$

is not satisfied,  $\alpha$  is decreased and a new coordinate vector  $(\phi_1, \phi_2, p_t)_{n+1}$  is computed. Thus, the  $n$ 'th stage of the process is repeated until Eq. 50 is satisfied or  $\alpha \leq \epsilon, \epsilon \ll 0$ .

- (2) Similarly,  $\alpha$  is decreased if Eq. 50 is satisfied during each of a successive number of steps.

- (3) The scaling parameters  $(s_j)$  are decreased each time a corresponding component of the gradient vector changes sign.

The process control philosophy is clearly an unsophisticated trial and error learning procedure. It does, however, provide a rapid and reliable method of handling the inevitable scaling problems associated with steepest descent or gradient methods. While it is true that more exact methods for determining the scaling matrix are available, such methods involve analytical or numerical evaluation of higher derivatives<sup>(10)</sup>. For the class of functions treated here, it is not clear that this additional sophistication is worth the cost (analytical and programming) of implementation. The fact that only a few seconds of IBM 7094 time is required to determine a typical local



minimum is offered as further testimony to the practicality of this simplified scaling procedure.

#### CONVERGENCE PROPERTIES

An impulse function space associated with a pair of inclined elliptical orbits, ( $p_1 = 5000$  mi,  $p_2 = 6000$  mi,  $e_1 = e_2 = 0.2$ ,  $\omega_1 = -90^\circ$ ,  $\omega_2 = +30^\circ$ ,  $i = 5^\circ$ ), was investigated with an IBM 7094 double precision program incorporating the adaptive steepest descent technique. This particular function space was previously studied in Ref. 1 by generating an optimum impulse contour map (Fig 1). The descent paths associated with a number of starting points have been plotted in Fig. 1. This function space offers no significant problems and the four minima predicted by contouring are quickly established with required accuracy (13 significant figures) regardless of the particular descent path. Table 1 contains the parameters associated with each minimum as well as the computer time required for the shorter descent paths.

Table 1 - Optimum Transfer Parameters

Initial Orbit $p_1 = 5000.$ mi $e_1 = 0.2$ $\omega_1 = -90.^\circ 0$ $i = 5.^\circ$					
Final Orbit $p_2 = 6000.$ mi $e_2 = 0.2$ $\omega_2 = +30.^\circ 0$					
Optimum	$\phi_1$ Deg.	$\phi_2$ , Deg.	$P_t$ , mi.	Impulse, fps.	7094 time, sec.
1	73.8152	187.5568	6644.8496	4902.65122 3852	3.4
2	40.8343	298.2634	6617.7904	5343.14869 3477	2.5
3	177.8114	73.6465	4611.8023	5393.78114 4757	3.0
4	308.2034	37.7403	4592.8574	5654.19120 9679	2.8

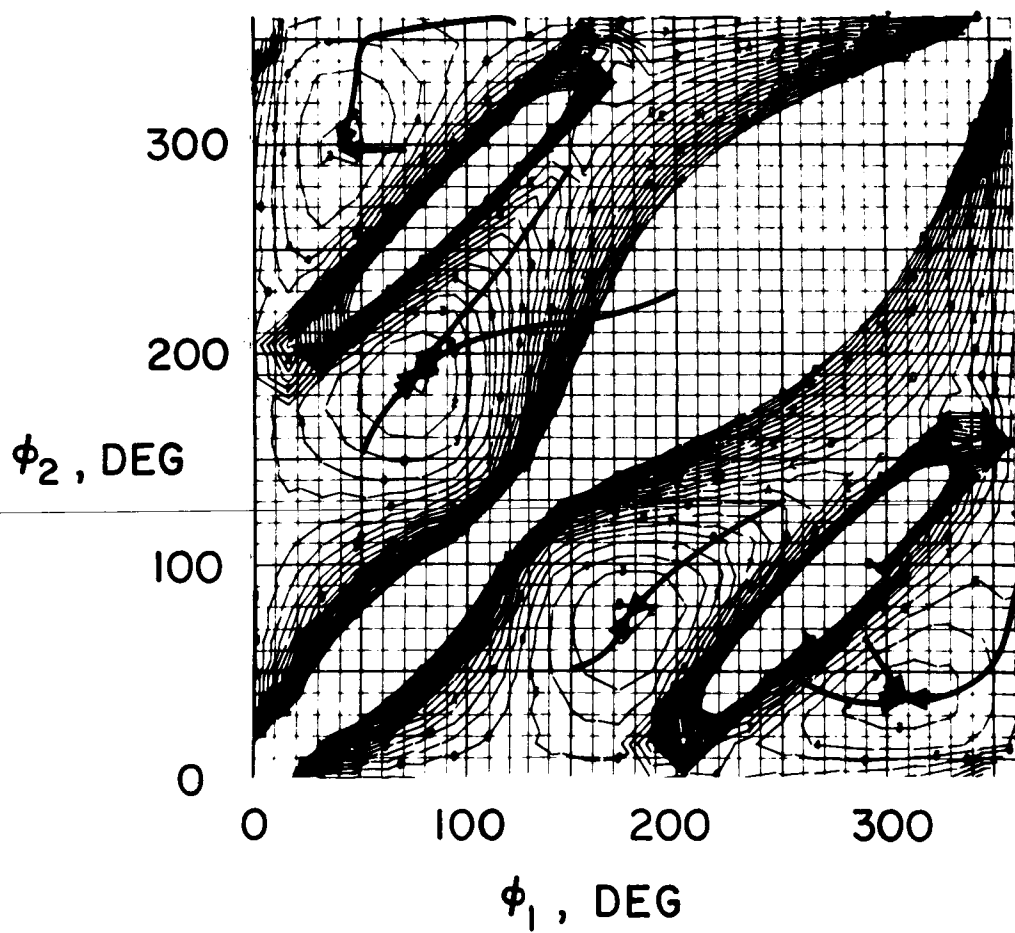


Figure 1 - Descent Paths Plotted on Optimum-Impulse Contour Map



Figs. 2 and 3 illustrate typical behavior of the various control and function parameters during optimization of a function having long narrow features similar to those appearing in Fig. 4. In order to minimize impulse in this example the program must search down a long narrow "tube" whose principal axis extends approximately in the  $\phi_1$  direction. The large initial increase in the  $\phi_1$  scale factor ( $s_1$ ) allows a large  $\phi_1$  correction to be accomplished early in the optimization (Fig. 2). Near the minimum the "tube" becomes more nearly "disc" shaped. The scaling parameters stabilize and maintain their general "disc" shape as the number of steps exceeds 500.

Fig. 3 illustrates convergence of the coordinate vector's components (solid lines). Note the large changes in  $\phi_1$ , corresponding to the maximum values of  $s_1$  appearing in Fig. 2. An additional case involving ordinary steepest descent optimization (i.e.,  $s_1 = s_2 = s_3 = 1$ ) is presented for comparison (broken lines). Under this constraint the descent process locates the center of the "tube" and then begins a very slow movement in the  $\phi_1$  direction. Impulse convergence for these two cases is also illustrated in Fig. 2. Note that the adaptive method continues to minimize long after the ordinary steepest descent process has essentially ceased optimization.

Although the convergence of the adaptive method appears to be quite slow, it should be remembered that this particular function was chosen for its difficulty. Fig. 2 also includes data for an optimization involving the inclined elliptical orbits which produced Fig. 1. For this optimization the impulse error ( $I_n - I_\infty$ ) decreases from  $10^4$  to  $10^{-5}$  in only 40 steps. Clearly, the method quickly adapts to the structure of any function and then proceeds to make good progress toward the local minimum of interest.

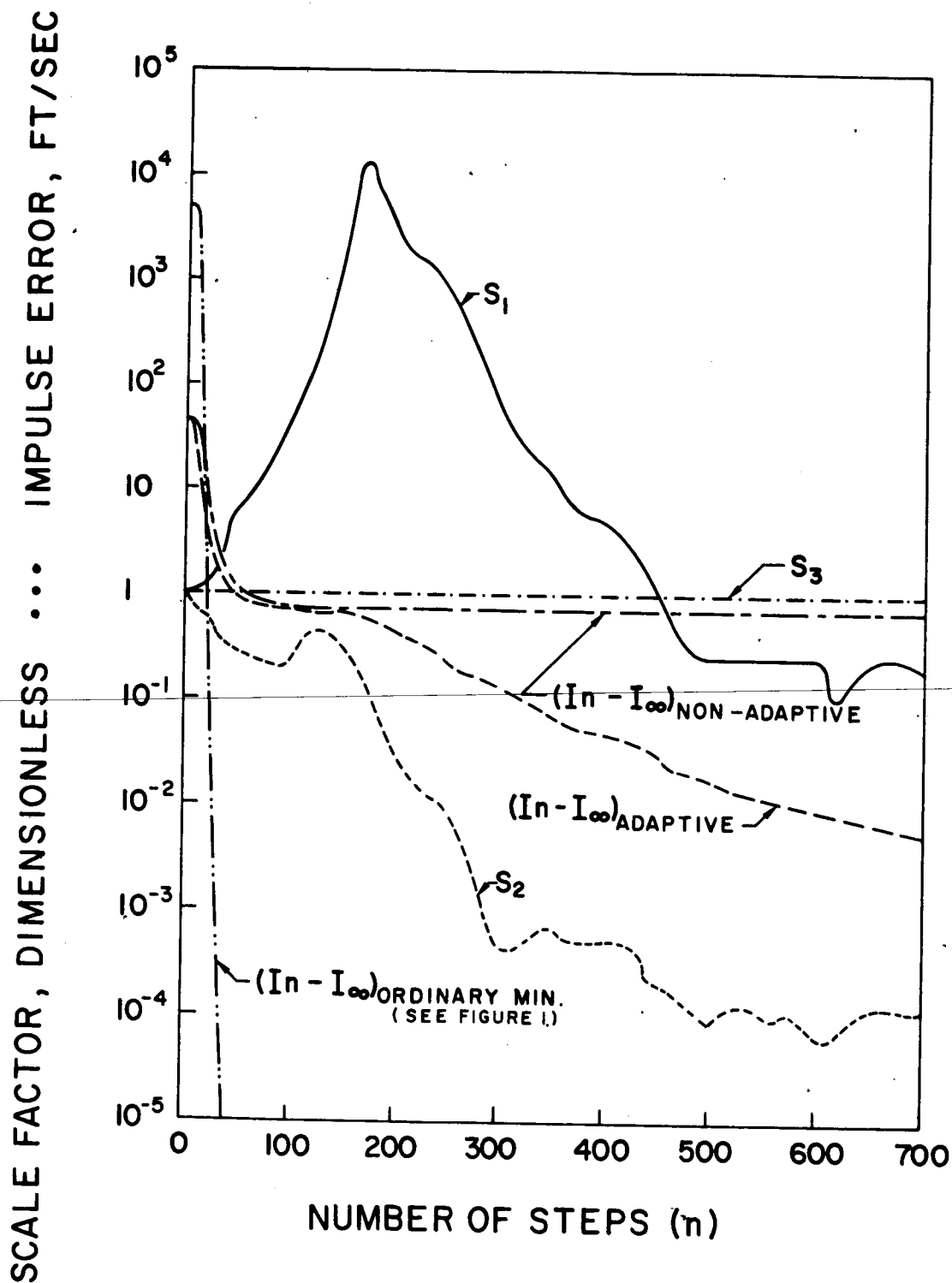


Figure 2 - Optimization Process (Parameter Variation)



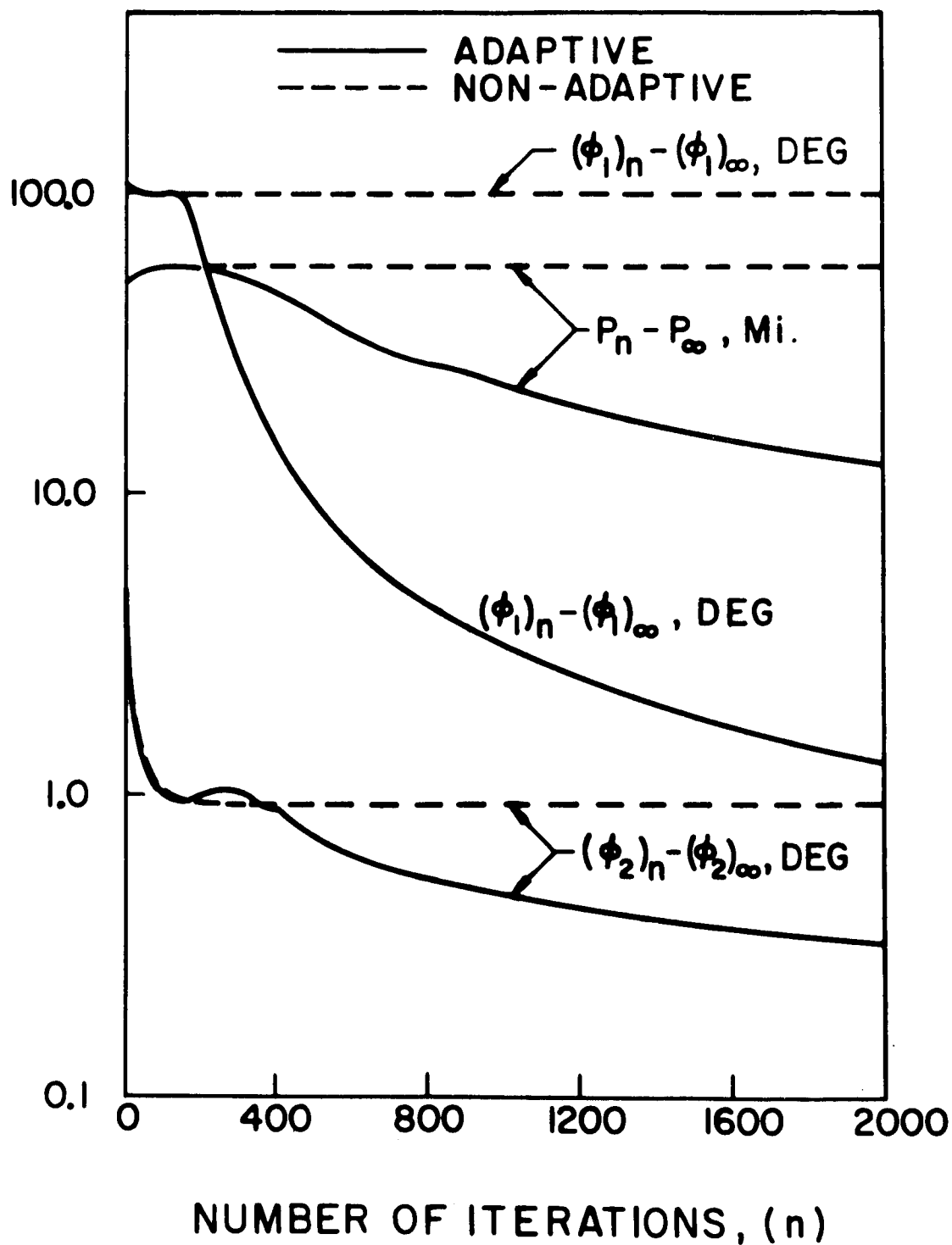


Figure 3 - Coordinate Vector Behavior During Optimization Process



## IV. NUMERICAL RESULTS

## "ALMOST TANGENT" ORBITS

Several authors have established the non-optimality of ordinary co-tangential transfers between elliptical orbits and one-impulse transfers at a point of tangency. (1, 11, 12, 13) However, one easily observes that optimum transfer orbits usually are nearly tangent to both the initial and final orbits. This fact and certain other questions generated during prior studies by function contouring (1, 2, 3) made the class of "almost tangent" orbits an interesting candidate for further numerical investigation. The existence of two locally optimum transfers between tangent orbits was demonstrated in Ref. 1. Further investigation using the adaptive steepest descent program has established the existence of at least three (3) local minima in the function spaces associated with a large class of "almost tangent" orbits.

In Fig. 4, two optimum impulse contour maps for a pair of tangent orbits ( $P_1 = 5000$  mi,  $P_2 = 6000$  mi,  $e_1 = e_2 = 0.2$ ,  $\Delta\omega = -53^\circ 13'01''$ ) are presented in order to adequately display the long narrow "valleys" which are characteristic of this class of function spaces. Note that the scales are greatly distorted to amplify certain details and to allow the use of a small contour interval ( $\Delta I = 0.01$  fps).

By constraining the numerical search to planes normal to the axes of the various valleys one may develop a complete picture of the optimum regions of a given function space. In Fig. 5 impulse is plotted as a function of position throughout the space by first traversing the horizontal valley ( $\phi_2 \approx 71^\circ$ ) and then traversing the vertical valley ( $\phi_1 \approx 71^\circ$ ). In Fig. 5b a number of points (a - f) are plotted on the curve for tangent orbits. These same points are reproduced in Fig. 4 to allow matching of the various

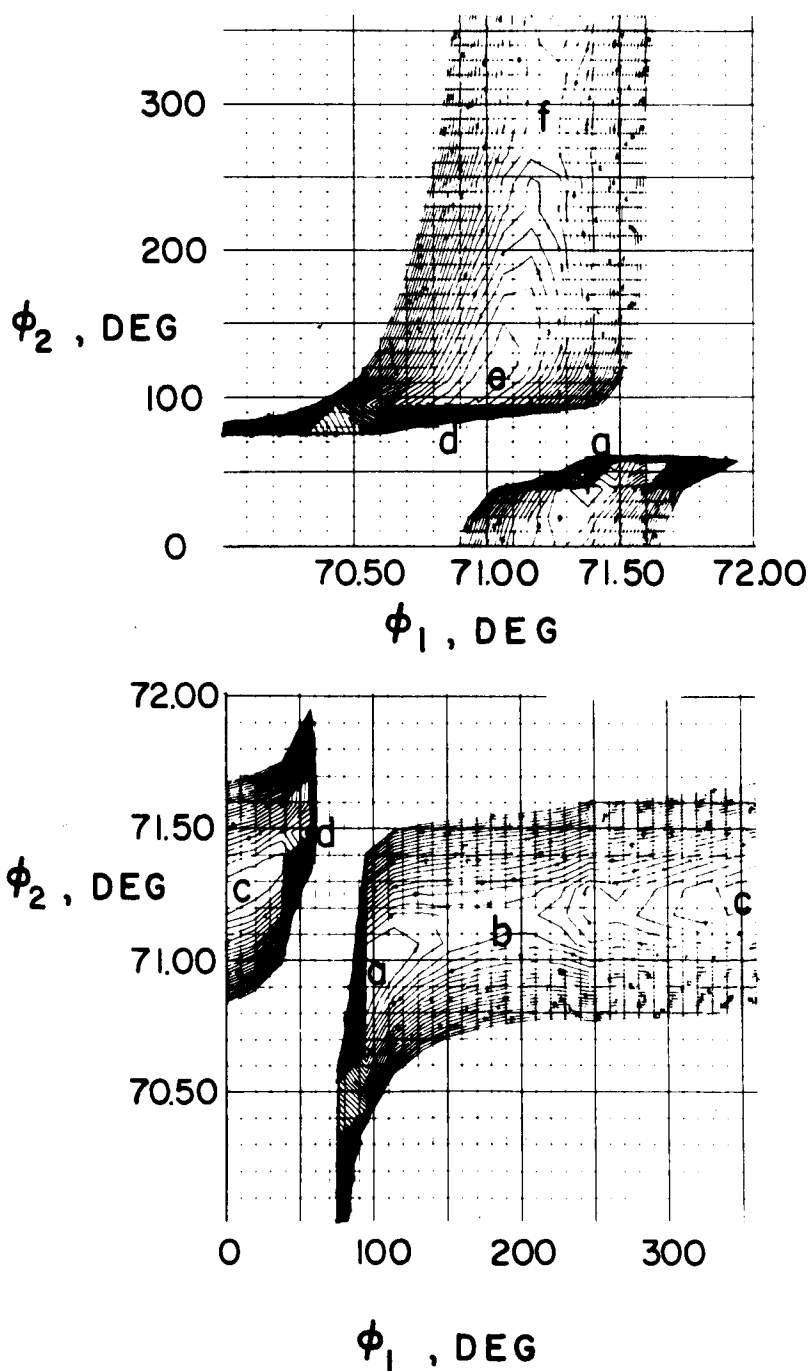
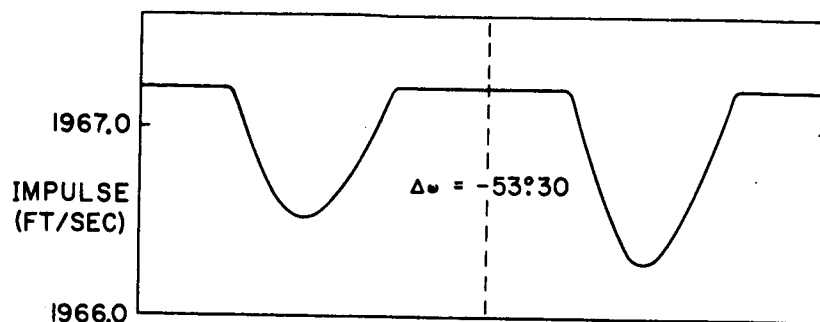
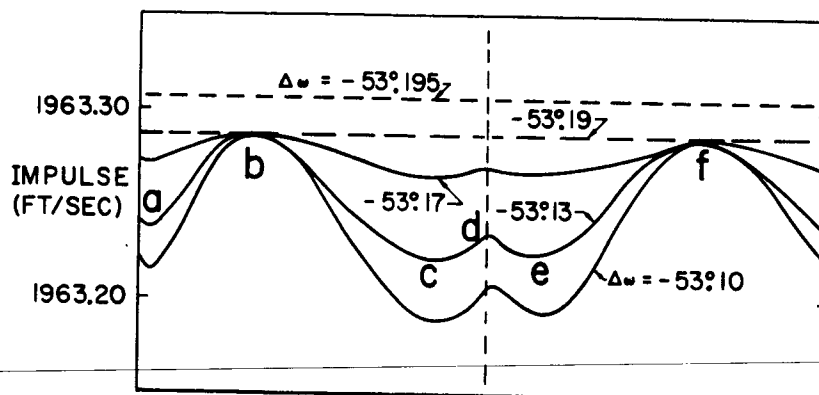


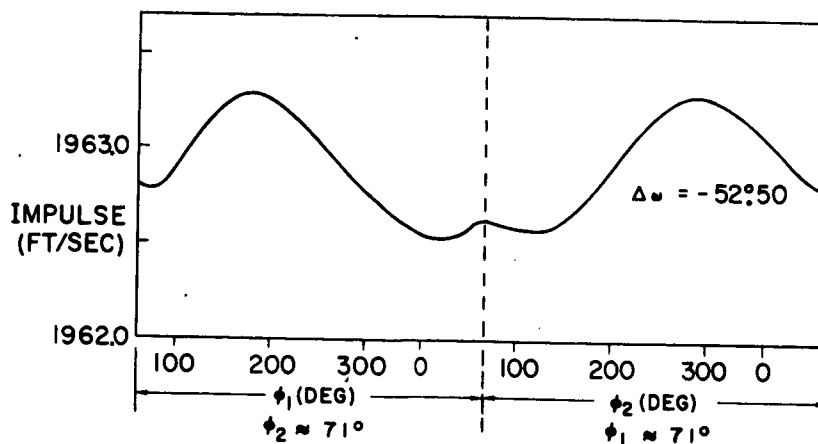
Figure 4 - Optimum Impulse Contour Maps for Tangent Orbits (contour interval = 0.01 fps.)



c) DEEPLY INTERSECTING ORBITS



b) "ALMOST TANGENT" ORBITS



a) NON-INTERSECTING ORBITS

Figure 5 - Optimum Impulse Variation Throughout Function Spaces - "Almost Tangent" Orbits



structural features with corresponding values of impulse.

The curves appearing in 5b were obtained by rotating the tangent orbits of Fig. 4 from a nonintersecting orientation ( $\Delta\omega = -53.^\circ 10$ ) to a slightly intersecting orientation ( $\Delta\omega = -53.^\circ 17$ ). Several of these curves exhibit three local minima. Although the impulse difference between minima is slight and not usually important in the engineering sense, it is necessary to isolate the absolute minimum for valid comparisons with finite thrust maneuvers such as the Lawden Spiral.

Also appearing in Fig. 5b are essentially straight lines corresponding to one-impulse transfer maneuvers performed at the intersection point of smallest radius. Contensou<sup>(15)</sup> and Breakwell<sup>(16)</sup> have each demonstrated the existence of such optimal one-impulse transfer maneuvers. The problem of finding these one-impulse maneuvers is discussed in Refs. 17 and 18 which develop formulae for predicting the range of orbit parameters for which the one-impulse maneuver is optimum. Figs. 5a and 5c illustrate the effect of large rotations from a tangency condition. Note that three local minima persist in Fig. 5a although the orbits are far from intersecting. If intersection deepens (Fig. 5c) the function space again begins to have small regions denoting two-impulse maneuvers which require less impulse than the associated one-impulse maneuver. Fig. 6 further clarifies this relationship by plotting optimum impulse for both the one and two-impulse maneuvers. The two curves are seen to coincide over a small range of relative orientation.

#### INCLINED ELLIPTICAL ORBITS

The existence of optimal coplanar orbital transfer maneuvers requiring no more than two impulses is discussed by Contensou<sup>(15)</sup> and

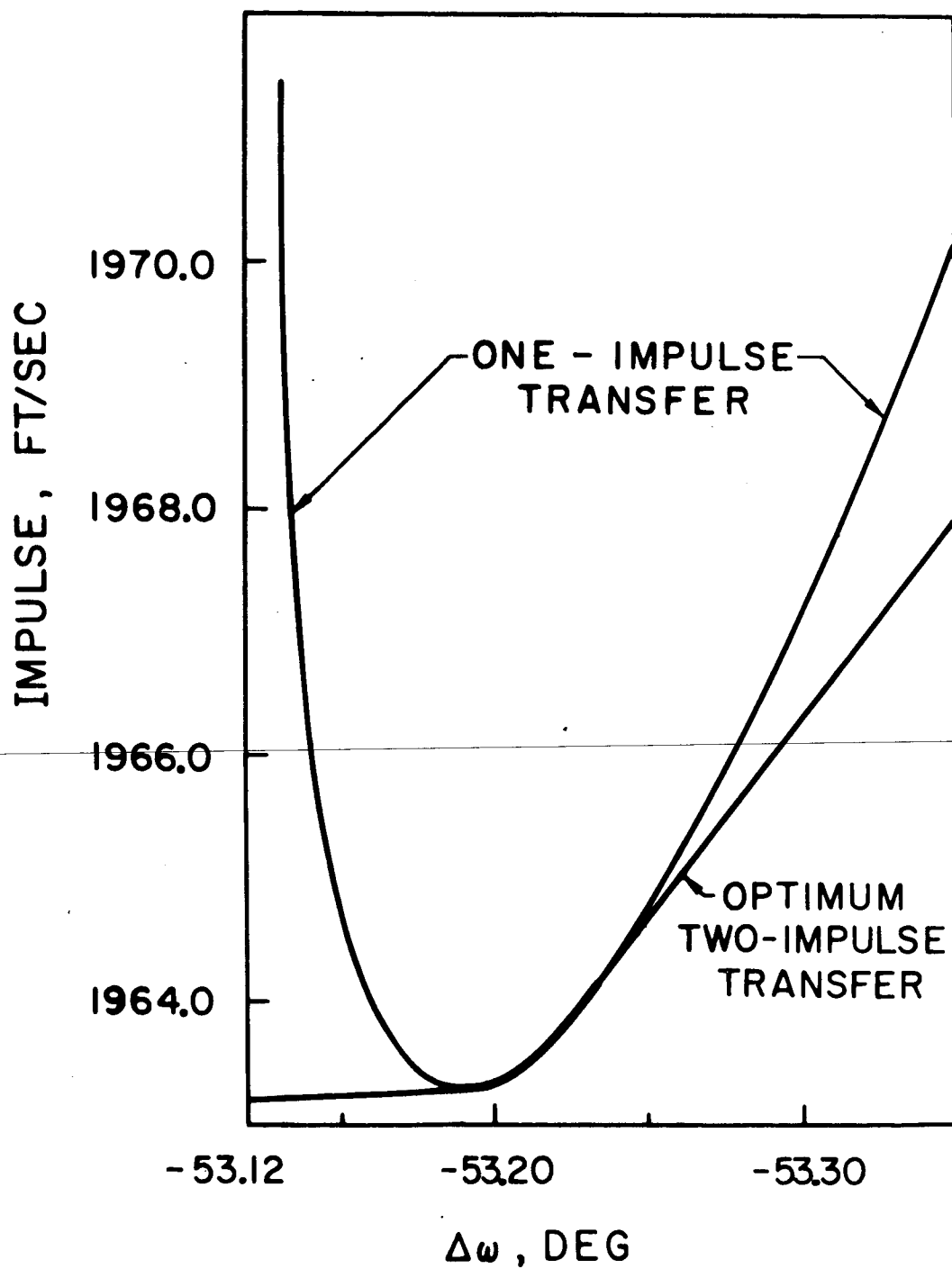


Figure 6 - Impulse Comparison for Optimum One and Two-Impulse Transfers



Breakwell. (16) Extensive investigation using the adaptive steepest descent program strongly suggests that optimal transfer between classes of inclined, non-coapsidal, elliptical orbits also requires a maximum of two impulses. It follows that optimal one-impulse maneuvers between inclined elliptical orbits must also exist.

#### LAWDEN SPIRAL VS. TWO IMPULSE TRANSFER

In Ref. 19 Lawden discusses the possible optimality of a particular intermediate thrust spiral trajectory. Using a contouring technique the authors of this paper demonstrated the existence of optimum two-impulse transfer maneuvers which require less total  $\Delta V$  than the corresponding Lawden spiral maneuvers (20, 21). Using the adaptive steepest descent program, these numerical results have now been expanded to give a broad comparison of the two-impulse maneuver and the Lawden spiral.

The orbits which oscillate to the Lawden spiral are generated by varying the parameter  $\sin^2 \psi$  which denotes position on the spiral. In Fig. 7 the difference in velocity change required for both maneuvers ( $\Delta V_{LS} - \Delta V_{2-imp}$ ) is plotted as a function of position difference between the osculation points. A family of curves was generated by varying  $\sin^2 \psi$  of the initial orbit.

In all cases computed a two-impulse maneuver which required less  $\Delta V$  than the Lawden spiral was found. Numerical accuracy limitations prevented extending these comparisons to smaller values of  $\Delta \sin^2 \psi$ . Interestingly enough, all the curves presented indicate that the difference in velocity change increases as the 4.7 power of  $\Delta \sin^2 \psi$ , which leads to a severe departure from the Lawden spiral  $\Delta V$  as  $\Delta \sin^2 \psi$  increases.

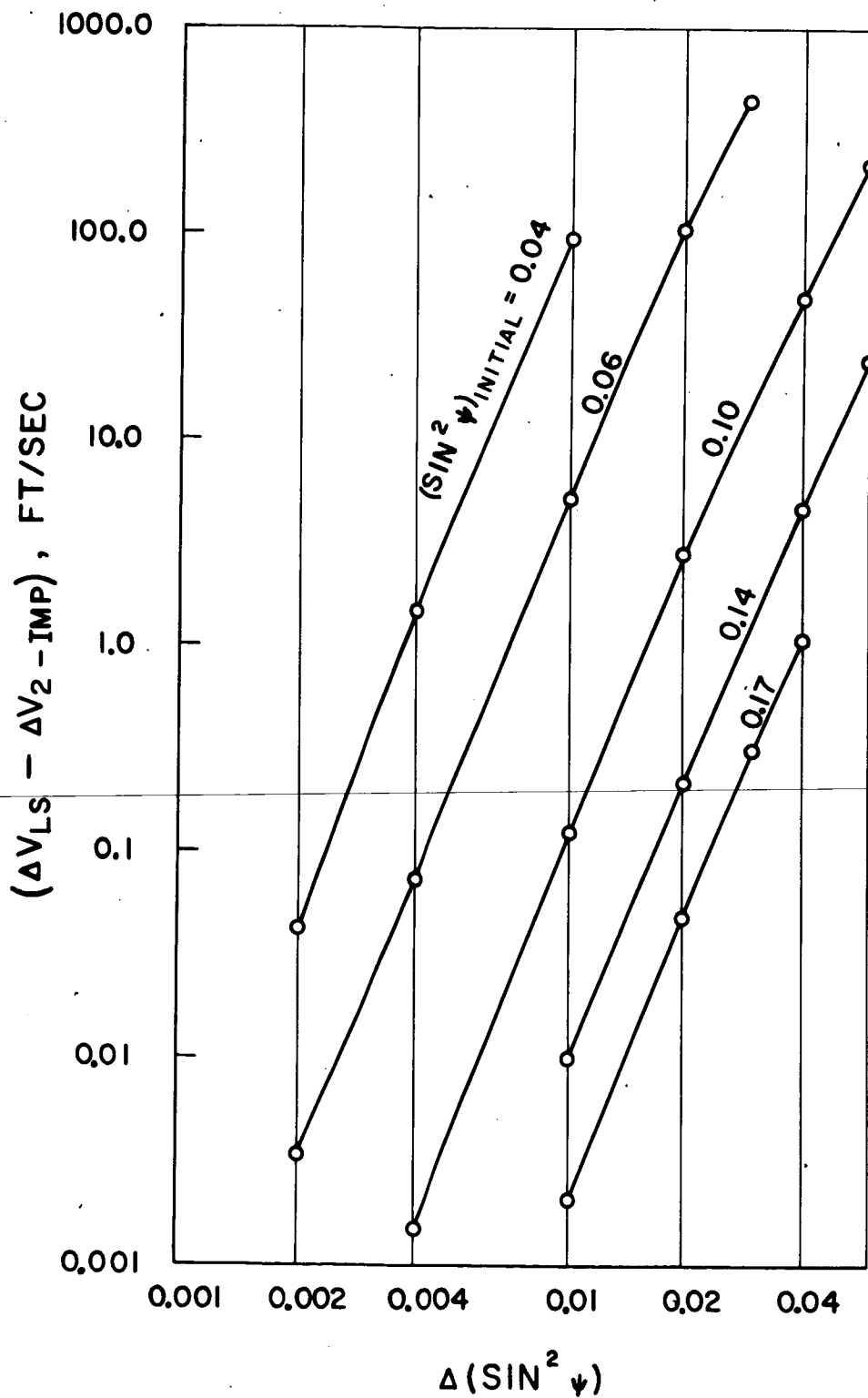


Figure 7 - Lawden Spiral  $\Delta V$  Compared to Optimum Two-Impulse  $\Delta V$





## V. CONCLUSION

An effective numerical method for precise computation of optimum two-impulse transfers between inclined elliptical orbits has been developed and verified. When supplemented by previously developed function mapping techniques, (1, 3) the adaptive steepest descent program has successfully minimized the most difficult function spaces encountered. The complexity of the more interesting function spaces suggests that considerable caution should be exercised when numerically seeking the absolute minimum two-impulse transfer.

In view of the demonstrated optimality of the two-impulse maneuver for transferring between a large class of orbits, this proven numerical optimization program becomes a valuable tool for use in numerous research and engineering studies.



## VI. REFERENCES

1. McCue, G. A., "Optimum Two-Impulse Orbital Transfer and Rendezvous Between Inclined Elliptical Orbits," AIAA J. 1, 1865-1872 (1963).
2. Des Jardins, P. R., Bender, D. F., and McCue, G. A., "Orbital Transfer and Satellite Rendezvous (Final Report)," SID 62-870, North American Aviation, Inc. (August 31, 1962).
3. McCue, G. A., "Optimization and Visualization of Functions," AIAA J. 2, 99-100 (1964).
4. Kerfoot, H. P., Bender D. F., and Des Jardins, P. R., "Analytical Study of Satellite Rendezvous (Final Report)," MD 59-272, North American Aviation, Inc., (October 20, 1960).
5. Lee, G., "An Analysis of Two-Impulse Orbital Transfer," Studies in the Fields of Space Flight and Guidance Theory, Progress Report No. 4, NASA MTP-AERO-63-65, 167-211 (1963).
6. Herget, P., "The Computation of Orbits," (published privately by the author, Ann Arbor, Michigan, 1948), p. 30.
7. Des Jardins, P. R., and Bender, D. F., "Extended Satellite Rendezvous Study," (Second Quarterly Report), SID 61-459, North American Aviation, Inc., (December 15, 1961).
8. Hoelker, R. F., "Orbit Transfer Studies by Numerical Processes," MTP-AERO-61-24, George C. Marshall Space Flight Center, (March 24, 1961).



9. Kerfoot, H. P., and Des Jardins, P. R., "Coplanar Two-Impulse Orbital Transfers," ARS Preprint 2063-61, (October 9, 1961).
10. Powell, M. J. D., "A Rapidly Convergent Descent Method for Minimization," The Computer Journal, 6, No. 2, 163-168, (July 1963).
11. Lawden, D. F., "Orbital Transfer via Tangential Ellipses," Journal of the British Interplanetary Society 2, No. 6, 278-289 (1952).
12. Bender, D. F., "Optimum Coplanar Two-Impulse Transfer Between Elliptic Orbits," J. of Aerospace Eng. 21, 44-52 (1962)
13. Li-Shu Wen, Wm., "A Study of Co-tangential, Elliptical Transfer Orbits in Space Flight," Journal of the Aerospace Sciences 28, No. 5 (1961).
14. Ting, L., "Optimum Orbital Transfer by Impulse," ARS J. 30, 1013-1018 (1960).
15. Contensou, P., "Etude Théorique Des Trajectoires Optimales Dans Un Champ De Gravitation. Application Au Cas D'Un Center D'Attraction Unique. (Theoretical Study of Optimal Trajectories in a Gravitational Field. Application in the Case of a Single Center of Attraction)," Astronautica Acta VIII, 2-3 (1963); also Grumman Res. Transl. Tr-22 by P. Kenneth (August 1962).
16. Breakwell, J. V., "Minimum Impulse Transfer," AIAA Paper 63-416.
17. Bender, D. F., and McCue, G. A., "Optimal One-Impulse Transfers Between Coplanar Elliptical Orbits," North American Aviation, Inc., (October 1964).
18. Bender, D. F., "A Comparison of One and Two-Impulse Transfer for Nearly Tangent Coplanar Elliptical Orbits," Studies in the Fields of Space Flight and Guidance Theory, Progress Report No. 5, NASA TM X-53024, 155-182 (March 17, 1964).



19. Lawden, D. F., "Optimal Intermediate-Thrust Arcs in a Gravitational Field," *Astronaut Acta* 8, 106-123.
20. Bender, D. F., and McCue, G. A., "Numerical Demonstration of the Non-Optimality of the Lawden Spiral," Presented at the 11th Technical Meeting Concerning Space Flight and Guidance Theory, MSFC (December 19, 1962).
21. Breakwell, J. V., "Unpublished Private Communication with Derek F. Lawden," Lockheed Missiles and Space Co., (December 18, 1962).

N65 33055

OPTIMAL ONE-IMPULSE TRANSFER  
BETWEEN COPLANAR ELLIPTICAL ORBITS

Prepared by

David F. Bender and Gary A. McCue

Space Sciences Laboratory  
Space and Information Systems Division  
North American Aviation, Inc.

Special Report No. 8

October 1, 1964

Contract NAS8-5211

---

Prepared for

George C. Marshall Space Flight Center  
National Aeronautics and Space Administration  
Huntsville, Alabama

**NORTH AMERICAN AVIATION, INC.**  
**SPACE and INFORMATION SYSTEMS DIVISION**

## TECHNICAL REPORT INDEX/ABSTRACT

ACCESSION NUMBER 69684-64						DOCUMENT SECURITY CLASSIFICATION Unclassified			
TITLE OF DOCUMENT OPTIMAL ONE-IMPULSE TRANSFER BETWEEN COPLANAR ELLIPTICAL ORBITS							LIBRARY USE ONLY		
AUTHOR(S) David F. Bender and Gary A. McCue									
CODE		ORIGINATING AGENCY AND OTHER SOURCES NAA - S&ID				DOCUMENT NUMBER SID NO. 64-1859			
PUBLICATION DATE October 1, 1964			CONTRACT NUMBER NAS 8-5211						
DESCRIPTIVE TERMS									

## ABSTRACT

Numerical and analytical results concerning optimum one-impulse orbital transfer maneuvers are presented. By considering a class of "shallowly intersecting" coplanar orbits which may be produced by differentially changing the orbital elements of a pair of tangent orbits, one may derive a number of approximate expressions concerning the minima of the one-impulse maneuvers that occur. Numerical comparisons of one-impulse transfers and corresponding optimum two-impulse and optimum  $180^\circ$  two-impulse transfers were made. These comparisons suggested that there exists a narrow range of shapes over which one-impulse transfer is optimal and indicated analytical expressions bounding the region for the equivalence of one-impulse transfer and optimum  $180^\circ$  two-impulse transfer. Simple exact equations which define outer bounds to the range of shapes over which one-impulse may be optimal were then derived. Straight forward evaluation of these expressions immediately establishes the non-optimality of a one-impulse transfer between any given pair of coplanar orbits.

Arthur



## CONTENTS

	Page
I. INTRODUCTION . . . . .	1
II. GEOMETRY OF SHALLOW INTERSECTIONS . . . . .	2
III. OPTIMAL ONE-IMPULSE TRANSFER . . . . .	5
IV. OPTIMIZED 180 DEGREE TWO-IMPULSE TRANSFER . . . . .	9
V. COMPARISON OF ONE AND TWO-IMPULSE TRANSFERS FOR "SHALLOWLY INTERSECTING" ORBITS . . . . .	11
VI. THE LIMITS FOR EQUIVALENCE OF ONE-IMPULSE AND OPTIMUM 180 DEG. TWO-IMPULSE TRANSFER . . . . .	17
VII. CONCLUSION . . . . .	24
VIII. REFERENCES . . . . .	25



## ILLUSTRATIONS

Figure		Page
1	The Geometry of Shallow Intersections	3
2	Impulse for One and Two-Impulse Transfers versus $\omega$ ( $\rho^2 = 1.2$ , $e_1 = 0.2$ , $e_2 = 0.2$ )	12
3	Impulse for One and Two-Impulse Transfers versus $\omega$ ( $\rho^2 = 1.8$ , $e_1 = 0.2$ , $e_2 = 0.6$ )	13
4	Impulse for One and Two-Impulse Transfers versus $\omega$ ( $\rho^2 = 2.25$ , $e_1 = 0.6$ , $e_2 = 0.95$ )	14
5	Optimum 180° Two-Impulse Transfers Versus Departure Point (Two-Impulse Optimal and Lower $\omega$ Limit)	18
6	Optimum 180° Two-Impulse Transfers Versus Departure Point (One-Impulse Optimal)	19
7	Optimum 180° Two-Impulse Transfers Versus Departure Point (Upper $\omega$ Limit and Two-Impulse Optimal)	20





## NOMENCLATURE

Scalars

a	Semimajor axis
e	Eccentricity
f	True anomaly
$\epsilon$	True anomaly half angle between intersection points
i	Inclination
j	Impulse
$\mu$	Gravitation constant
p	Semilatus rectum
$\rho$	$\sqrt{p_2/p_1}$
$\phi$	Angle denoting true anomaly of line which bisects angle between intersection points

---

$\omega$	Argument of perigee (initial orbit relative to final orbit)
----------	---

Vectors

$\underline{P}$	Unit vector in perigee direction
$\underline{Q}$	Unit vector 90° ahead of perigee vector
$\underline{V}$	Velocity vector

Subscripts

T	Tangent
1	Initial orbit parameter, or intersection point of smallest radius
2	Final orbit parameter, or intersection point of largest radius
3	180° two-impulse transfer parameter
m	Minimum



## I. INTRODUCTION

During the course of a continuing study of optimum orbital transfer maneuvers (Refs. 1, 2, 3, 4, 5) the class of "shallowly intersecting" orbit pairs was shown to be worthy of further study. For such orbits, numerical data indicated the existence of one-impulse orbital transfer maneuvers which resulted in minimum fuel expenditure; a result which has been discussed by Contensou<sup>(6)</sup>, and Breakwell<sup>(7)</sup>. If one must find the optimum transfer between a pair of non-coapsidal, "shallowly intersecting," coplanar elliptical orbits, it is clearly desirable to determine if a one-impulse maneuver is optimal before proceeding with two-impulse optimization techniques such as those described in Refs. 4 and 5. Furthermore, it is known that the impulse function spaces associated with such orbit pairs offer a formidable and time consuming challenge to numerical optimization techniques.<sup>(5)</sup> This is largely because these function spaces are structured in the form of long narrow "valleys" containing several minima.<sup>(5)</sup> Therefore, a strong motivation for developing formulae for predicting and evaluating these favorable orbital transfer maneuvers exists.



## II. GEOMETRY OF SHALLOW INTERSECTIONS

Consider two coplanar, non-coapsidal, elliptical orbits that are nearly tangent and are described by the elements:  $p_1, p_2 = \rho^2 p_1$  where  $\rho^2 = p_2/p_1 > 1$ ,  $e_1 \neq 0$ ,  $e_2 \neq 0$ ,  $\omega_2 = 0$ , and  $\omega_1 = \omega \neq 0$ . For  $p_1 = p_2$  the orbit intersections must lie 180 degrees apart and this case is therefore excluded because a shallow intersection is to be characterized by a small true anomaly interval between the two points of intersection. Finally, one may restrict  $\omega$  to the range,  $0 < \omega < 180^\circ$  without loss of generality since only the angular difference between the perigee vectors ( $\underline{P}_1$  and  $\underline{P}_2$ ) is required.

The geometry of the shallow intersection is shown in Fig. 1. Let the line FB lie at the angle  $\phi$  from  $\underline{P}_1$ , and let it also bisect the angle between the two intersections ( $2\epsilon$ ). If  $\epsilon$  is small, and,  $0 < 2\epsilon < 180^\circ$ :

$$\sin \phi = e_2 \sin \omega/D \quad (1)$$

$$\cos \phi = (e_2 \cos \omega - \rho^2 e_1)/D \quad (2)$$

$$\cos \epsilon = (\rho^2 - 1)/D \quad (3)$$

where,

$$D^2 = \rho^4 e_1^2 + e_2^2 - 2 \rho^2 e_1 e_2 \cos \omega \quad (4)$$

The true anomalies of the intersection points, of smallest and largest radius are  $\phi - \epsilon$  and  $\phi + \epsilon$  respectively. Let the subscript "T" denote tangent orbits and assume that the elements  $p$ ,  $e_1$ ,  $e_2$ , and  $\omega$  differ by small amounts ( $\delta p$ ,  $\delta e_1$ ,  $\delta e_2$ ,  $\delta \omega$ ) from their values at tangency, and furthermore, assume that these perturbed orbits intersect. Since the tangent condition

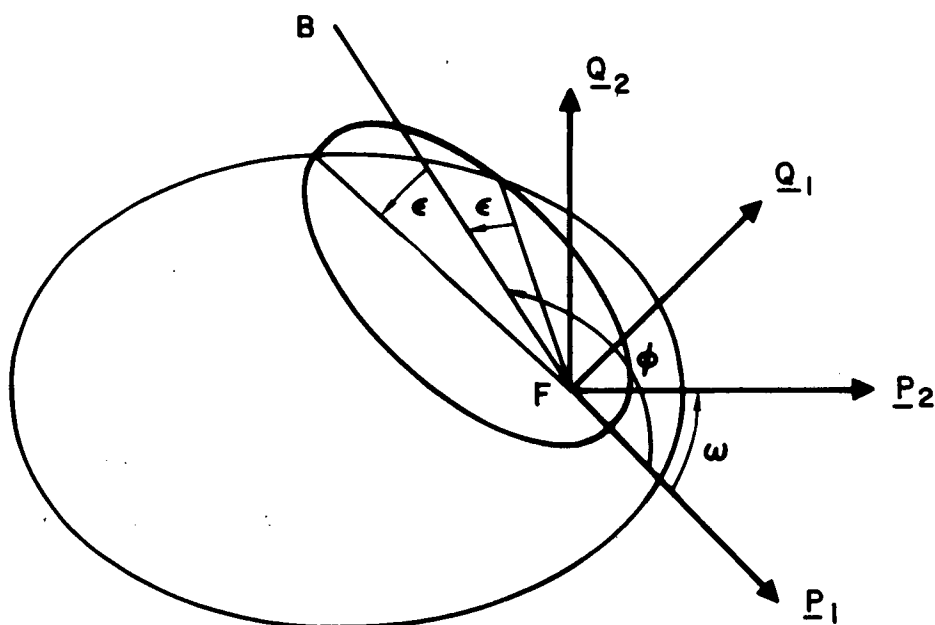


Figure 1 - The Geometry of Shallow Intersections



requires that:

$$D_T^2 = (\rho^2 - 1)^2 = \rho^4 e_1^2 - 2 \rho^2 e_1 e_2 \cos \omega + e_2^2 \quad (5)$$

for  $\epsilon \ll 1$  one may write

$$1 - \cos \epsilon \approx \frac{\epsilon^2}{2} \approx \sum_{j=1}^4 \frac{\partial (\cos \epsilon)}{\partial \alpha_j} \delta \alpha_j \quad (6)$$

where  $\alpha_j$ ; are the four elements:  $\rho$ ,  $e_1$ ,  $e_2$ , and  $\omega$ .

Clearly, Eq. 5 may be used to find the value of an element that will yield tangency if the other elements are given.

$$\text{Thus,} \quad \epsilon \approx \sqrt{-2 \sum_{j=1}^4 \frac{\partial \cos \epsilon}{\partial \alpha_j} \delta \alpha_j} \quad (7)$$

Although shallow intersection may be generated by differentially changing any of the four parameters, only changes in  $\omega$  will be considered for brevity in the numerical comparisons. For small changes,  $\delta \omega$ , one may assume  $\omega_T = \omega$  and,

$$D^2 - D_T^2 = 2 \rho^2 e_1 e_2 [\cos \omega_T - \cos (\omega_T + \delta \omega)] \quad (8)$$

$$\approx 2 \rho^2 e_1 e_2 \sin \omega \delta \omega \quad (9)$$

and

$$\epsilon \approx \frac{\rho}{\rho^2 - 1} \sqrt{2 e_1 e_2 \sin \omega \delta \omega} \quad (10)$$



### III. OPTIMAL ONE-IMPULSE TRANSFER

One and two-impulse transfers between coplanar orbits have been investigated by Ting<sup>(8)</sup>, Horner<sup>(9)</sup>, and Barrar<sup>(10)</sup>. Although these authors did not specifically consider nearly tangent orbits Barrar does mention the possibility of optimizing one-impulse transfer on orbit orientation

It is convenient to adopt the notation of Ref. 11, and to express velocities and impulses in units of  $\sqrt{\mu/p_1}$ . Velocity vectors ( $\underline{v}_1$  and  $\underline{v}_2$ ) in the initial and final orbits may then be defined as follows:

$$\sqrt{\frac{p_1}{\mu}} \underline{v}_1 = \underline{V} + e_1 \underline{Q}_1 \quad (11)$$

$$\sqrt{\frac{p_1}{\mu}} \underline{v}_2 = \frac{1}{\rho} (\underline{V} + e_2 \underline{Q}_2) \quad (12)$$

where  $\underline{V}$  is a unit vector perpendicular to the radius at the transfer point and  $\underline{Q}_1$  and  $\underline{Q}_2$  are unit vectors perpendicular to the perigee vectors (See Fig. 1). The impulse for the one-impulse transfer maneuver, is expressed as follows:

$$\underline{j} = \frac{\underline{V} (1 - \rho) + \underline{C}}{\rho} \quad (13)$$

where

$$\underline{C} = e_2 \underline{Q}_2 - \rho e_1 \underline{Q}_1 \quad (14)$$

$$j^2 \rho^2 = C^2 + (1 - \rho)^2 + 2 (1 - \rho) \underline{C} \cdot \underline{V} \quad (15)$$

$$C^2 = \underline{C} \cdot \underline{C} = \rho^2 e_1^2 - 2 \rho e_1 e_2 \cos \omega + e_2^2 \quad (16)$$



and

$$\underline{C} \cdot \underline{V} = e_2 \cos (\phi - \epsilon - \omega) - \rho e_1 \cos (\phi - \epsilon) \quad (17)$$

where the angles  $(\phi - \epsilon)$  and  $(\phi - \epsilon - \omega)$  are the true anomalies of the transfer point on the first and second orbits respectively. By using Eqs. 1 and 2 the angle  $\phi$  is eliminated and Eq. 17 becomes:

$$\underline{C} \cdot \underline{V} = \frac{\cos \epsilon}{D} E^2 + \frac{\sin \epsilon}{D} (\rho - 1) \rho e_1 e_2 \sin \omega \quad (18)$$

where

$$E^2 = \rho^3 e_1^2 - (1 + \rho) \rho e_1 e_2 \cos \omega + e_2^2 \quad (19)$$

Finally,

$$j^2 \rho^2 = C^2 + (\rho - 1)^2 - 2(\rho - 1) \left[ \frac{E^2}{D} \cos \epsilon + \frac{(\rho - 1) \rho e_1 e_2 \sin \omega \sin \epsilon}{D} \right] \quad (20)$$

One can now compare the impulse for the tangent condition,  $(j_T)$ , with the impulses for the two points of intersection  $(j_1 \text{ and } j_2)$  by changing the sign of  $\epsilon$  in Eq. 20. Clearly,  $\epsilon$  must be small as must  $\delta \rho$ ,  $\delta e_1$ ,  $\delta e_2$ , and  $\delta \omega$ .

Noting that

$$j^2 - j_T^2 = (j - j_T)(j + j_T) \approx 2j_T(j - j_T) \quad (21)$$

it follows that

$$\begin{aligned} j - j_T \approx \frac{1}{2j_T} & \left[ \left( \frac{C^2}{\rho^2} - \frac{C_T^2}{\rho_T^2} \right) + \left( \frac{(\rho - 1)^2}{\rho^2} - \frac{(\rho_T - 1)^2}{\rho_T^2} \right) \right. \\ & + \left( - \frac{2(\rho - 1) E^2}{D \rho^2} \cos \epsilon + \frac{2(\rho_T - 1) E_T^2}{D_T \rho_T^2} \right) \\ & \left. - \frac{2(\rho - 1)^2 e_1 e_2 \sin \omega}{D} \epsilon \right] \quad (22) \end{aligned}$$



If each of the paired terms in Eq. 22 is expressed as a Taylor series about the tangent condition the leading coefficients will involve  $\delta\rho$ ,  $\delta e_1$ ,  $\delta e_2$ , or  $\delta\omega$ . However, the term involving  $\epsilon$  has in its leading coefficient  $\sqrt{\delta\rho}$ ,  $\sqrt{\delta e_1}$ ,  $\sqrt{\delta e_2}$ , or  $\sqrt{\delta\omega}$  (see Eq. 7). Therefore, as long as it does not have a zero coefficient the latter will dominate the expression as small changes are introduced. Since  $\epsilon$  can be positive or negative it follows that for one intersection the impulse to transfer is at first less than that required at tangency, and for the other intersection it is greater. The intersection corresponding to  $-\epsilon$  is the one with the smaller impulse and smaller radius—a result pointed out by Anthony and Sasaki. (12)

If  $\rho$ ,  $e_1$  and  $e_2$  are fixed and only  $\omega$  is varied, Eq. 22 yields:

$$j_1 - j_T \approx \frac{e_1 e_2 \sin \omega}{j_T} \left[ \frac{2 E^2 \delta \omega}{(\rho + 1) D^2} + \frac{(\rho - 1)^2 \epsilon}{\rho D} \right] \quad (23)$$

Removing  $\epsilon$  by using Eq. 10 gives (for  $\epsilon$  positive):

$$j_1 - j_T \approx \frac{e_1 e_2 \sin \omega}{j_T (\rho + 1)} \left[ 2 \frac{E^2}{D^2} \delta \omega - \frac{\sqrt{2 e_1 e_2 \sin \omega} \delta \omega}{\rho + 1} \right] \quad (24)$$

The terms neglected in Eq. 24 begin with  $\delta\omega^{3/2}$ , and  $\delta\omega$  has to be positive in the direction which yields the pair of shallow intersection. Since the sign of the coefficient of  $\delta\omega$  is positive Eq. 24 has a minimum which is given by Eq. 25.

$$(\delta\omega)_m = \frac{e_1 e_2 \sin \omega}{8(\rho + 1)^2 E^4 / D^4} \quad (25)$$

The corresponding values of  $(\epsilon)_m$  and impulse change relative to tangency are:

$$(\epsilon)_m = \frac{\rho}{\rho^2 - 1} \frac{e_1 e_2 \sin \omega}{2(\rho + 1) E^2 / D^2} \quad (26)$$





## IV. OPTIMIZED 180 DEGREE TWO-IMPULSE TRANSFER

Determining an optimum two-impulse transfer is in general a three parameter problem wherein even the conditions for optimum transfer between coplanar orbits yield extremely unwieldy expressions. By contrast, finding optimum 180 deg. two-impulse transfer is a two parameter problem and optimization of one of the parameters is easily accomplished. In addition, numerical comparisons<sup>(13)</sup> indicate that optimum 180 deg. transfers closely approximate the optimum two-impulse transfer in many cases. For these reasons, and because simplified expressions are available for use in later derivations, certain equations for 180° transfer are presented here.

Considering optimization of the transfer orbit parameter, the departure point being fixed but arbitrary, the impulse is given by: (with minor modifications to the notation of Ref. 13).

$$j_3 = \sqrt{(x_1 - x_2)^2 + (y_1 + y_2)^2} \quad (28)$$

where

$$x_1 = e_1 \sin f \quad (29)$$

$$x_2 = \frac{e_2 \sin (f - \omega)}{\rho} \quad (30)$$

$$y_1 = (1 + e_1 \cos f) |c_3 - 1| \quad (31)$$

$$y_2 = \frac{1 - e_2 \cos (f - \omega)}{\rho^2} | \rho - c_3 | \quad (32)$$



and,

$$(j_1 - j_T)_m = - \frac{e_1^2 e_2^2 \sin^2 \omega}{4j_T (\rho + 1)^3 E^2/D^2} \quad (27)$$



where,  $f$  = true anomaly of the departure point on the first orbit,

$$C_3^2 = \frac{p_3}{p_1} = 2 / \left\{ 1 + e_1 \cos f + [1 - e_2 \cos (f - \omega)] / \rho^2 \right\} \quad (33)$$

$j_3$  = total impulse required in units of  $\sqrt{\frac{\mu}{p_1}}$ ,

and,  $p_3$  = semilatus rectum of transfer orbit.



## V. COMPARISON OF ONE AND TWO-IMPULSE TRANSFERS FOR "SHALLOWLY INTERSECTING" ORBITS

Numerical results were obtained by first determining the intersection of shortest radius and then searching for the optimum 180 deg. two-impulse transfer by varying the departure point. A search was initiated by determining what is called a practically optimum transfer in Ref. 13. Numerical investigations of numerous orbit pairs all yielded similar results. For purpose of illustration three orbit pairs with very different values of eccentricity are presented here: 1)  $\rho^2 = 1.2$ ,  $e_1 = e_2 = 0.2$ ,  $\omega_T = \cos^{-1} 0.6 = 53.^\circ 1301$ ; 2)  $\rho^2 = 1.8$ ,  $e_1 = 0.2$ ,  $e_2 = 0.6$ ,  $\omega_T = 110.3741^\circ$ ; and 3)  $\rho^2 = 2.25$ ,  $e_1 = 0.6$ ,  $e_2 = 0.95$ ,  $\omega_T = 63.0498^\circ$ .

One-impulse and optimum 180 deg. two-impulse transfer data is shown in Figs. 2, 3, and 4. For this example the intersection producing element variation was obtained by rotating the final orbit relative to the initial orbit. The two-impulse curves are seen to coincide with the one-impulse curves near the minimum, the differences between the two being in the computer noise (8 decimal places) over a small but finite range of relative orientation. A few points on the two-impulse curve were investigated by a fully optimized double precision two-impulse program (Ref. 5) and these points are indicated by the black dots in Fig. 2. For these illustrations no significant difference between optimum two-impulse transfers and optimum 180 deg. transfer is apparent.

An intersection-producing change in shape can also be generated by varying  $e_1$  (or  $e_2$ ) or  $\rho$ . Computer studies of such cases yielded curves similar to those of Figures 2, 3, and 4. In every case the one-impulse

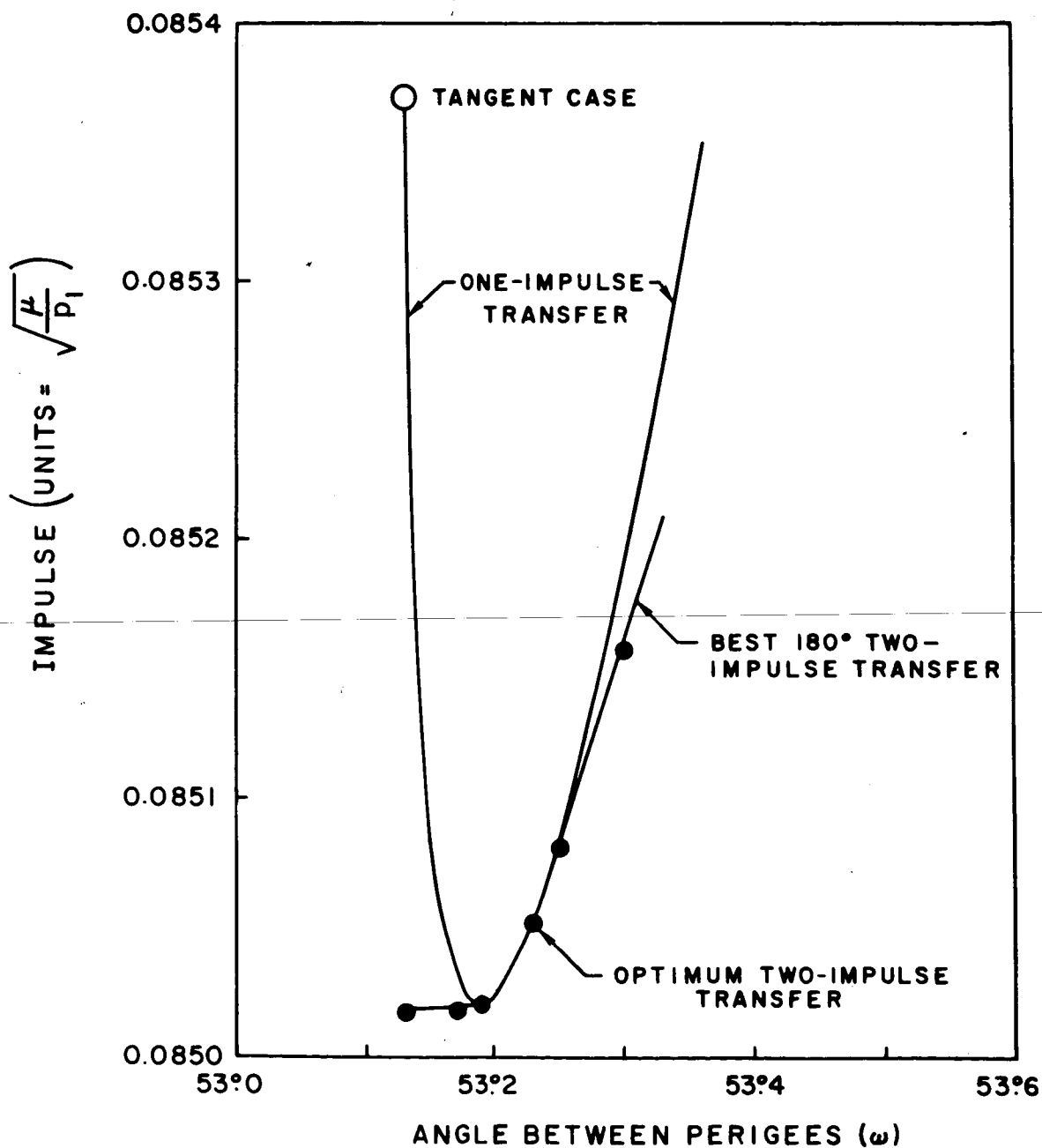


Figure 2 - Impulse for One and Two-Impulse Transfers versus  $\omega$   
 ( $p^2 = 1.2$ ,  $e_1 = 0.2$ ,  $e_2 = 0.2$ )

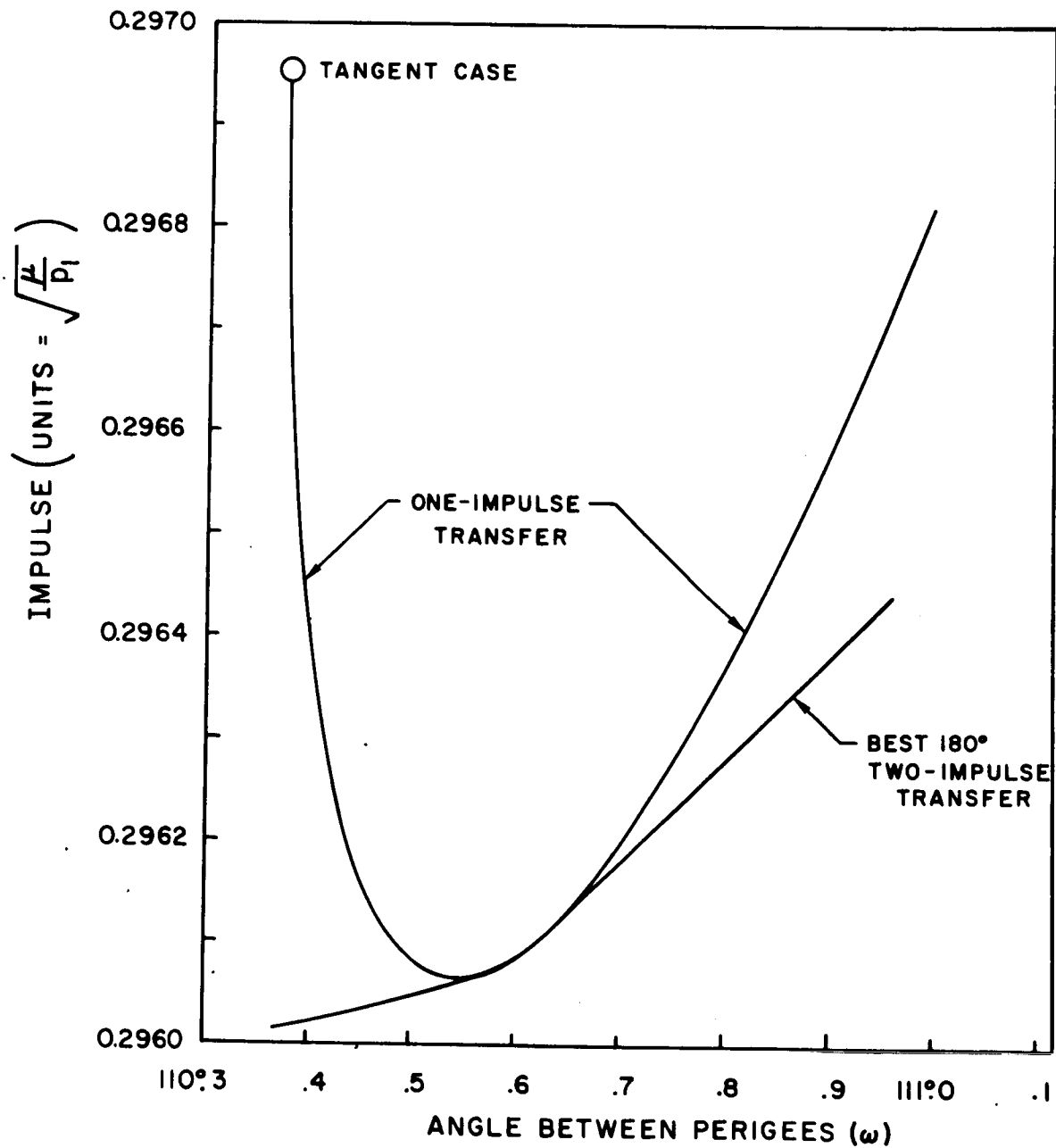


Figure 3 - Impulse for One and Two-Impulse Transfers versus  $\omega$   
 ( $p^2 = 1.8$ ,  $e_1 = 0.2$ ,  $e_2 = 0.6$ )

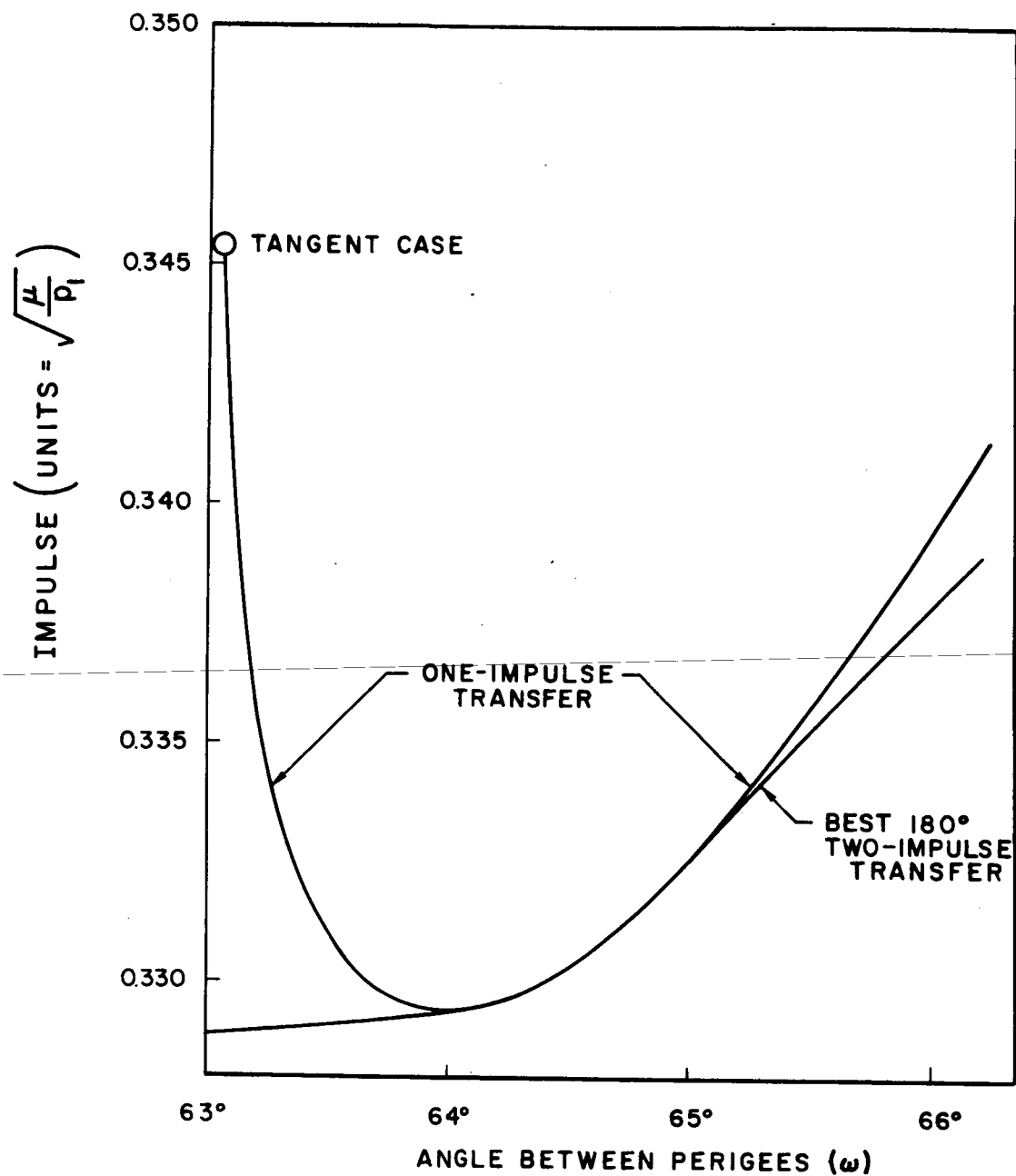


Figure 4 - Impulse for One and Two-Impulse Transfers versus  $\omega$   
( $p^2 = 2.25$ ,  $e_1 = 0.6$ ,  $e_2 = 0.95$ )



transfer experiences a minimum near tangency, and in every case the two-impulse curve coincides with this minimum as it does for the cases presented.

Table 1 summarizes the results of using the approximate formulae to predict values of  $(\delta\omega)_m$ ,  $(\epsilon)_m$ ,  $(j_T - j_1)_m$  for the three cases illustrated. The values indicated "(pred.)" were obtained from Eqs. 25, 26, and 27 while the values labeled "(comp.)" were obtained by a one-impulse computer program. The predicted values show good agreement with the actual values; even for the highly eccentric orbits.





TABLE 1. PARAMETERS DESCRIBING OPTIMAL ONE AND TWO-IMPULSE TRANSFERS NEAR TANGENCY

Case	Fixed Elements	$\omega_2 - \omega_1$ , Deg (Tangency) (Min. 1-Imp.)	Impulse (Tangency) (Min. 1-Imp.)	$(\delta\omega)_m$ , deg (pred) (comp)	$(\epsilon)_m$ , deg (pred) (comp)	$(j_T - j_1)_m$ (pred) (comp)	$\omega$ Limits for Optimal One-Impulse Transfer, deg.
1.	$\rho^2 = 1.2$	53.1301	.08536856	.059	2.56	.000348	53.1892 to 53.2330
	$e_1 = e_2 = .2$	53.1895	.08502114 (1963.28)*	.059	2.56	.000348 (8.03)*	
2.	$\rho^2 = 1.8$	110.3741	.2969501	.174	2.51	.00087	110.5467 to 110.6346
	$e_1 = .2$ $e_2 = .6$	110.545	.2960641 (6241.0)**	.171	2.60	.00089 (18.9)**	
3.	$\rho^2 = 2.25$	63.0498	.3454466	1.076	9.504	.01626	64.1042 to 64.6224
	$e_1 = .6$ $e_2 = .95$	64.042	.3293834 (5172.1)***	.992	9.057	.01606 (255.5)***	

\* In ft/sec for  $P_1 = 5000$  miles\*\* In ft/sec for  $P_1 = 6000$  miles\*\*\* In ft/sec for  $P_1 = 10,800$  miles



# VI. THE LIMITS FOR EQUIVALENCE OF ONE-IMPULSE AND OPTIMUM 180 DEG. TWO-IMPULSE TRANSFER

Figs. 5, 6, and 7 present a sequence of curves for optimum 180 deg. two-impulse transfer in the region where the two curves are identical. The first pair of coplanar elliptical orbits ( $\rho^2 = 1.2$ ,  $e_1 = 0.2$ ,  $e_2 = 0.2$ ) is involved. The single impulse transfer is always at the point of discontinuity and it is to be noted that the curves include this particular transfer whether or not it happens to be optimum. Each graph consists of two curves: one for which the departure point is near the intersection point and the first impulse is large (labeled "+" and referring to the positive scale of departure points) and one for which the arrival point is near the intersection and the second impulse is large (labeled "-" and referring to the negative scale of departure points). It is thus seen that the range of values of  $\omega$  over which the best 180° two-impulse transfer reduces to the single impulse at intersection can be indicated by requiring the proper curve to exhibit horizontal tangent as the intersection is approached from the proper side. (Note the scale changes which were required to plot the various small differences.) Nearly similar sets of curves were obtained for other pairs of orbits but are not shown. Of course, one may also cause the set of shapes to be given by a range of values of  $e_1$  or  $\rho^2$  instead of  $\omega$ . Again a similar set of curves would be obtained, indicating a range of values of the variable over which one-impulse transfer at intersection is identical with the best two-impulse transfer (180° two-impulse transfer).

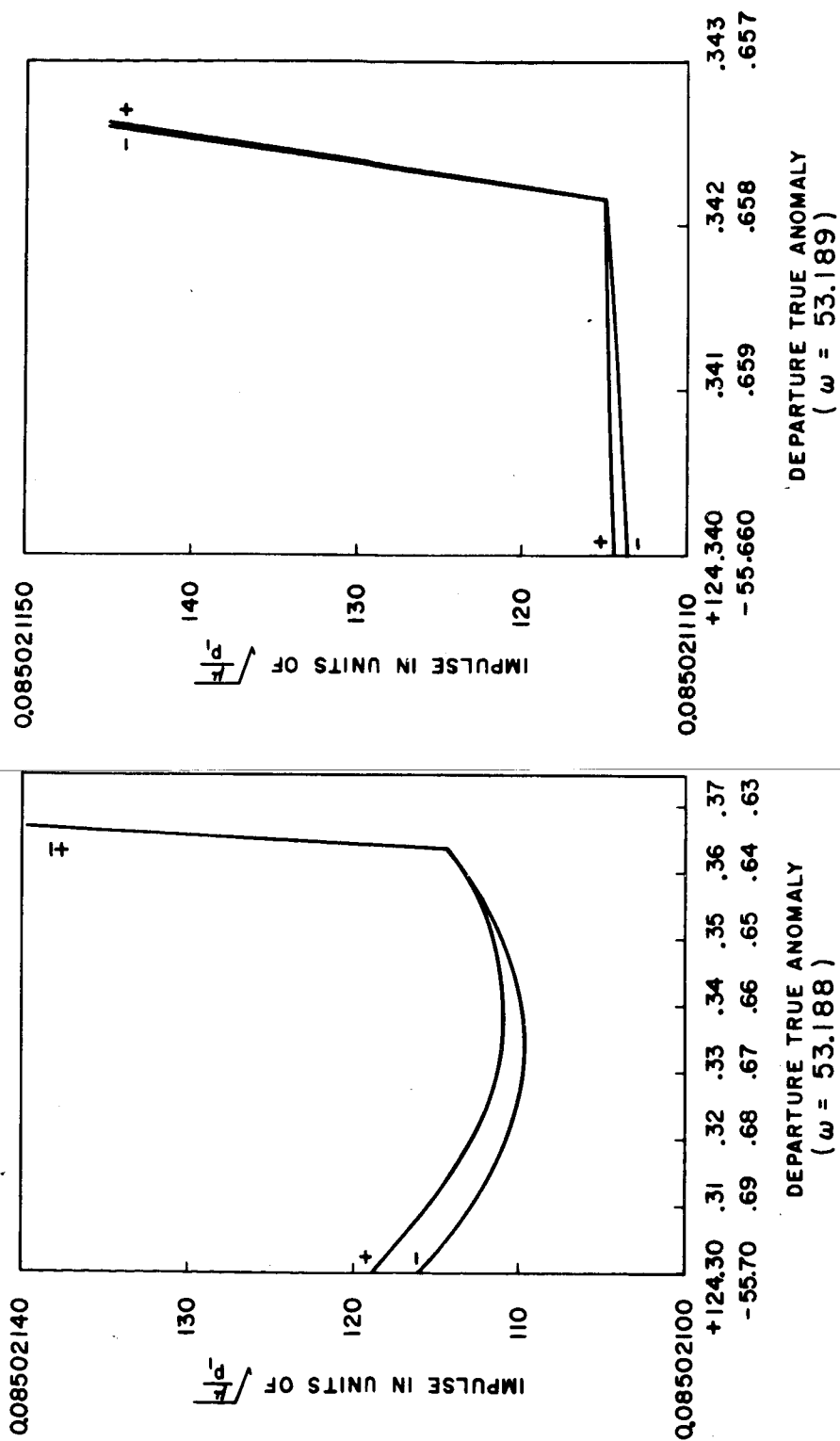


Figure 5 - Optimum 180° Two-Impulse Transfers Versus Departure Point (Two-Impulse Optimal and Lower  $\omega$  Limit)

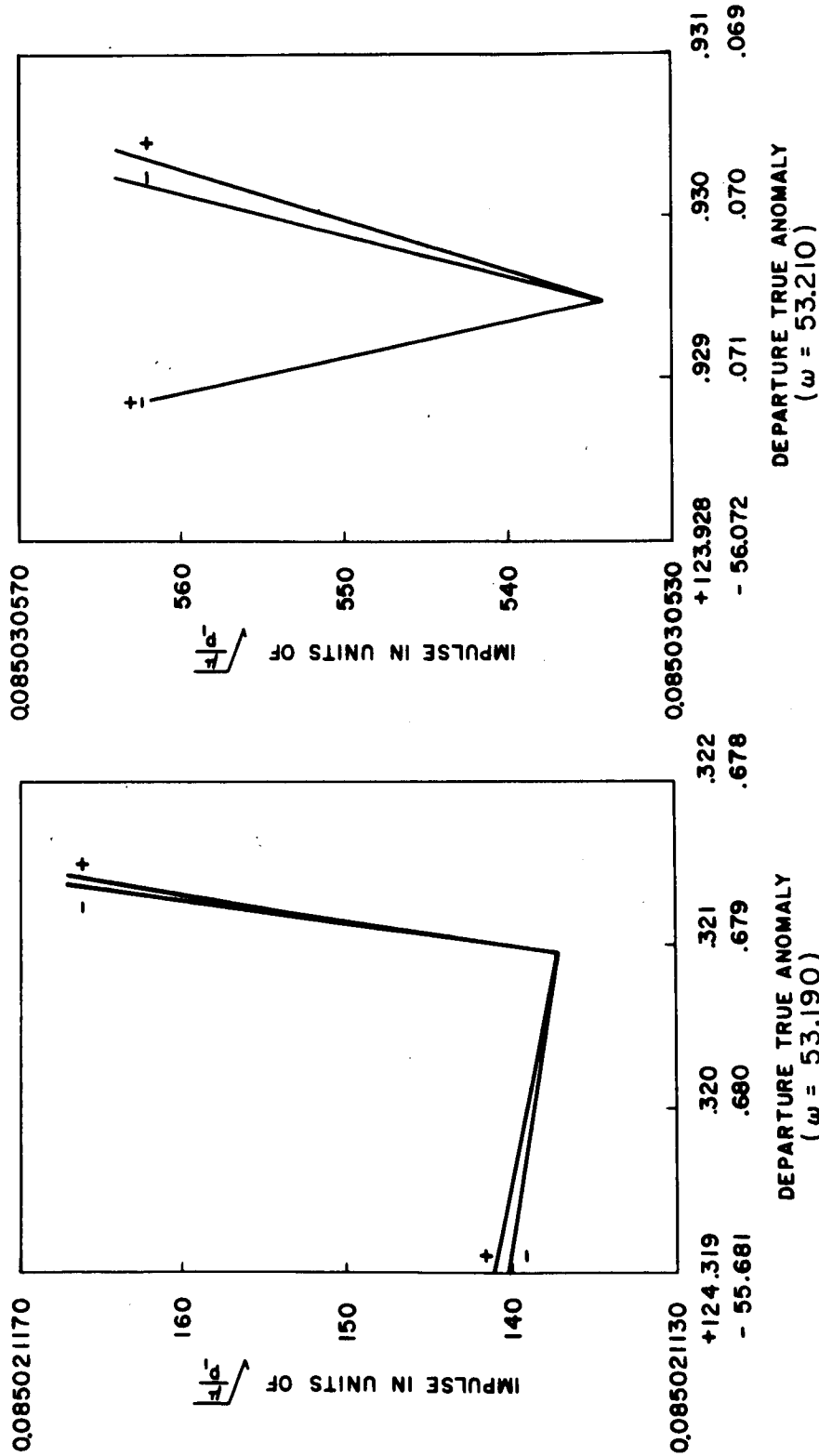


Figure 6 - Optimum  $180^\circ$  Two-Impulse Transfers versus Departure Point  
(One-Impulse Optimal)

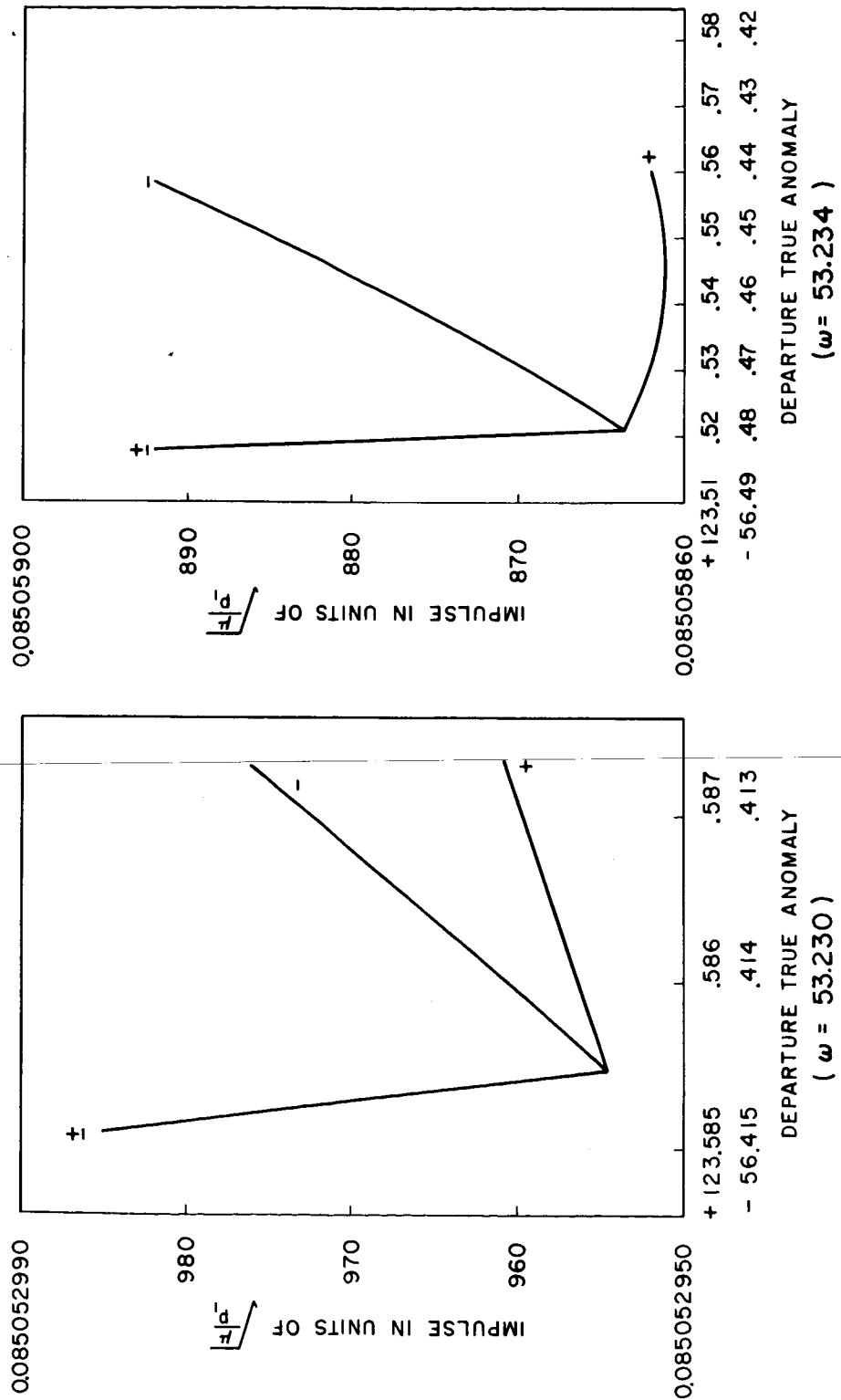


Figure 7 - Optimum 180° Two-Impulse Transfers versus Departure Point  
(Upper  $\omega$  Limit and Two-Impulse Optimal)



For all curves on the left of the point of intersection the small impulse is in the forward direction since the angular momentum (proportional to  $p_3^{\frac{1}{2}}$ ) lies between that of the initial and final orbits while for all curves on the right of the intersection point the small impulse opposes the direction of motion. In addition, for every case the + curve is above the - curve on the left and below on the right. The following condition, therefore, bounds the optimum one-impulse transfer region:

(a) upper limit; + curve has a horizontal tangent on the right or,

$$\left( \frac{dj_3^2}{df} \right)_{f = f_1 +} = 0, \quad \text{with } C_3 = \rho + \quad (34)$$

b) lower limit; - curve has a horizontal tangent on the left or,

$$\left( \frac{dj_3^2}{df} \right)_{f = (180^\circ + f_1) -} = 0, \quad \text{with } C_3 = 1. + \quad (35)$$

To express the limit conditions determined in Section V in a more manageable form, it is necessary to evaluate the slope ( $dC_3/df$ ) and use the proper sign. In addition, restrict  $f$  to  $f = \emptyset - \epsilon$  in all equations and in the frequently used expression.

$$p_1/r_1 = 1 + e_1 \cos f = \frac{1 + e_2 \cos (f - \omega)}{\rho^2} \quad (36)$$



The equations become

$$\left( e_1 \sin f - \frac{e_2 \sin (f - \omega)}{\rho} \right) \left[ e_1 \cos f + \frac{\rho + 1}{\rho} \right] \\ + (1 + e_1 \cos f) (\rho - 1) e_1 \sin f =$$

$$1) \quad \frac{D \sin \epsilon}{2\rho} (1 + e_1 \cos f) \quad (\text{upper limit}) \quad (37)$$

$$2) \quad \frac{D \sin \epsilon}{2\rho^2} (1 + e_1 \cos f) (2\rho - 2 + e_1 \cos f) \quad (\text{lower limit}) \quad (38)$$

Replacing  $f$  by  $\phi - \epsilon$ , one obtains equations involving  $\sin \epsilon$ , and  $\cos \epsilon$  up to the third power. By substituting  $\cos \epsilon = \frac{\rho^2 - 1}{D}$ , and  $\sin^2 \epsilon = 1 - \cos^2 \epsilon$ , one obtains equations which are linear in  $\sin \epsilon$  and may be solved easily.

$$\sin \epsilon = \frac{e_1 e_2 \sin \omega}{D} \frac{2(\rho^2 - 1)(\rho - 1) - D^2 + (\rho^2 - 1)^2}{G^2} \quad (\text{upper limit}) \quad (39)$$

$$= \left\{ \frac{e_1 e_2 \sin \omega}{D} \right\} \left\{ 2(\rho^2 - 1)(\rho - 1) - \right. \\ \left. \left[ D^2 - (\rho^2 - 1)^2 \right] \left[ \frac{1}{\rho} - 2e_1(\rho^2 - 1) \cos \phi / \rho D^2 \right] \right\} \\ \left\{ G^2 - \frac{1}{\rho} \left[ D^2(\rho - 2 + e_1^2 \sin^2 \phi \right. \right. \\ \left. \left. + D e_1(\rho^2 - 1)(\rho - 1) \cos \phi \right. \right. \\ \left. \left. + e_1^2(\rho^2 - 1)^2 \cos 2\phi \right] \right\}^{-1} \quad (\text{lower limit}) \quad (40)$$

where,

$$G^2 = 3\rho^2 e_1^2 - (\rho^2 + 2\rho + 3) e_1 e_2 \cos \omega + \left( 1 + \frac{2}{\rho} \right) e_2^2$$

A simple iteration scheme was programmed to solve Eqs. 39 and 40. When checked by using a double precision program the limiting values for  $\omega$  corresponded to the expected horizontal tangents in the graphs of impulse versus departure point.



In this development the variations in shape were made to occur as a result of variation in relative perigee angle. The results (Eqs. 39 and 40) however, are quite general. Thus in order to determine whether or not one-impulse transfer may be the optimum for any given pair of intersecting orbits one would evaluate  $\epsilon$  from  $\cos \epsilon = (\rho^2 - 1)/D$  and then determine whether or not  $\sin \epsilon$  lies in the range  $\sin \epsilon_1$  to  $\sin \epsilon_u$ . If so, one-impulse transfer at the intersection with the smaller radius may be the optimum impulse transfer between the two orbits.





## VII. CONCLUSION

A series of formulae for investigating the existence and properties of optimal one-impulse transfers between pairs of "shallowly intersecting" elliptical orbits have been developed and verified. In all cases tested, each of two different two-impulse optimization programs converged upon optimum one-impulse transfers predicted by these formulae. Numerical experiments with orbit pairs obtained using Breakwell's procedure <sup>(7)</sup> have shown that such orbits satisfy the conditions specified by Eqs. 37 and 40. As a result one may now discover these optimal maneuvers before proceed with two-impulse optimization procedures such as those described in Refs. 4 and 5.

---



## REFERENCES

1. Kerfoot, H. P., and Des Jardins, P. R., "Coplanar Two-Impulse Orbital Transfers," ARS Preprint 2063-61, (October 9, 1961).
2. Kerfoot, H. P., Bender, D. F., and Des Jardins, P. R., "Analytical Study of Satellite Rendezvous (Final Report)," MD 59-272, North American Aviation, Inc., (October 20, 1960).
3. Des Jardins, P. R., Bender, D. F., and McCue, G. A., "Orbital Transfer and Satellite Rendezvous (Final Report)," SID 62-870, North American Aviation, Inc. (August 31, 1962).
4. McCue, G. A., "Optimum Two-Impulse Orbital Transfer and Rendezvous Between Inclined Elliptical Orbits," AIAA J. 1, 1865-1872 (1963).
5. McCue, G. A., and Bender, D. F., "Numerical Investigation of Minimum Impulse Orbital Transfer," North American Aviation, Inc. SID 64-423 (October 1, 1964).
6. Contensou, P., "Etude Théorique Des Trajectories Optimales Dans Un Champ De Gravitation. Application Au Cas D'Un Center D'Attraction Unique. (Theoretical Study of Optimal Trajectories in a Gravitational Field. Application in the Case of a Single Center of Attraction)," Astronautica Acta VIII, 2-3 (1963); also Grumman Res. Transl. TR-22 by P. Kenneth (August 1962).
7. (Breakwell, J. V., "Minimum Impulse Transfer," AIAA Paper 63-416.
8. Ting, L., "Optimum Orbital Transfer by Impulse," ARS J. 30, 1013-1018 (1960).



9. Horner, J., "Minimum Impulse Orbital Transfer," AIAA J. 1, 1707-1708 (1963).
10. Barrar, R. B., "Two-Impulse Transfer vs. One-Impulse Transfer: Analytic Theory," AIAA J. 1, 65-68 (1963).
11. Bender, D. F., "A Comparison of One and Two-Impulse Transfer for Nearly Tangent Coplanar Elliptical Orbits," Studies in the Fields of Space Flight and Guidance Theory, Progress Report No. 5, NASA TM X-53024, 155-182 (March 17, 1964).
12. Anthony, M. L., and Sasaki, F. T., On Some Single Impulse Transfer Problems, AIAA Paper 63-421. (1963a)
13. Bender, D. F., "Optimum Coplanar Two-Impulse Transfers Between Elliptic Orbits," Aerospace Eng. 21, 10, 44-52.

# Research Laboratories

**U**  
**UNITED AIRCRAFT CORPORATION**  
**A**  
**EAST HARTFORD, CONNECTICUT**

Report C-910098-12

Optimal Variable-Thrust Rendezvous  
of a Power-Limited Rocket Between  
Neighboring Low-Eccentricity Orbits

Contract NAS8-11099  
Final Report

REPORTED BY

*F. W. Gobetz*  
F. W. Gobetz

N65  
33056

APPROVED BY

*J. L. Cooley*  
J. L. Cooley

Chief, Evaluation Section

DATE October 1964

NO. OF PAGES 65

COPY NO. \_\_\_\_\_

Report C-910098-12

Optimal Variable-Thrust Rendezvous of a Power-Limited Rocket  
Between Neighboring Low-Eccentricity Orbits

Contract NAS8-11099

Final Report

TABLE OF CONTENTS

	<u>Page</u>
SUMMARY.....	1
CONCLUSIONS.....	1
RECOMMENDATIONS.....	2
INTRODUCTION.....	2
ANALYTICAL METHOD	
Description of the Mathematical Model.....	3
Analysis.....	5
Synthesis of the Optimal Controls.....	6
RESULTS	
Orbit Transfer and Rendezvous.....	7
Choice of Reference Orbit.....	8
Application to Planetary Orbits.....	8
Extension of the Linearized Theory.....	9
Control Synthesis.....	9
REFERENCES.....	11
LIST OF SYMBOLS.....	12
APPENDIXES	
I - ROTATING RECTANGULAR AND SPHERICAL COORDINATE SYSTEM.....	15

TABLE OF CONTENTS  
(contd.)

	<u>Page</u>
II - LAGRANGE'S VARIABLES.....	22
III - SYNTHESIS OF THE OPTIMAL CONTROLS.....	34
FIGURES.....	40

Optimal Variable-Thrust Rendezvous of a Power-Limited Rocket

Between Neighboring Low-Eccentricity Orbits

Contract NAS8-11099

Final Report

SUMMARY

39056

A study has been made of minimum-fuel transfer and rendezvous between neighboring low-eccentricity orbits by power-limited rocket. This study includes and extends previous work wherein only the case of transfer between circular orbits was considered. As before, the analysis is based on the assumption that only small deviations from an initial orbit are allowed. Complete analytical solutions are obtained in three different sets of variables: (1) rotating rectangular coordinates, (2) rotating spherical coordinates, and (3) Lagrange's planetary variables. In addition to the determination of optimal transfer and rendezvous trajectories in three dimensions, synthesis of the optimal controls is also carried out in each case. The guidance coefficients resulting from the control synthesis are presented both in graphical form and in equation form suitable for use in guidance applications.

The use of an intermediate reference orbit is found to be a powerful method of improving the accuracy of the linearized theory. Results for circular, coplanar earth-Venus and earth-Mars transfers are compared with exact solutions. The linear theory is shown to provide a very good correlation with exact data for all trip times of interest.

*Author*

CONCLUSIONS

1. Explicit solutions are obtainable for minimum-fuel transfer and rendezvous between neighboring low-eccentricity orbits by power-limited rockets. These solutions include closed form expressions for the optimum thrust vector, the optimum trajectory, and the minimum required fuel consumption in terms of boundary conditions and trip time.

2. Synthesis of the optimal control has also been carried out for both transfer and rendezvous between any orbit and a neighboring, low-eccentricity orbit. Guidance coefficients for each case can be presented in terms of time remaining to reach the target orbit.

3. Results for the case of coplanar circle-to-circle transfer between earth and Venus indicate that the linearized equations adequately predict the actual motion, the optimal control, and the minimum fuel consumption. There is, as yet, no numerical data to indicate that the rendezvous equations are equally applicable to the planetary orbits. The failure of these equations appears to be caused by the terms representing the angular motion.

#### RECOMMENDATIONS

The results of the linearized analysis for earth-Mars and earth-Venus transfers are sufficiently promising to warrant further investigation into higher-order theories. In particular, the "piecewise-linear" theory described herein is a relatively straightforward application of the linearized equations which should include at least some second-order effects on the motion. It is recommended that this approach be pursued because a simple second-order solution is highly desirable.

#### INTRODUCTION

It is characteristic of high-specific-impulse, low-thrust propulsion systems that the source of power is separate from the thrust device itself. Consequently, such propulsion systems are referred to as power-limited, since thrust is restricted in magnitude by the output of the power supply, which is in turn limited by the necessity of minimizing power supply weight.

The problem of transfer and rendezvous between neighboring orbits by a power-limited rocket is of interest for two basic reasons. First of all, the problem can be solved analytically, as was demonstrated in Refs. 1, 2, and 3, provided that the thrust acceleration is not constrained in magnitude and that the proper simplifying assumptions are made in the mathematical model of the system. The analytic expressions thus obtained for the controls and for the optimum trajectories then provide insight into more general problems where the simplifying restrictions are lifted. Secondly, the solution to this problem provides a lower bound to the performance requirements for low-thrust orbital transfer and rendezvous.

It is interesting to note that if, for the same system model as has been used herein, the thrust acceleration is assumed constant, analytic integration



of the equations of motion requires the evaluation of incomplete elliptic integrals of the third kind (Ref. 4). Therefore, allowance for variable-thrust acceleration is essential if simple analytic solutions are to be obtained.

## ANALYTICAL METHOD

### Description of the Mathematical Model

The phrase "neighboring orbits", as defined here, requires that the inclination between orbit planes be small and that the radial separation between orbits be small relative to the semi-major axis of either orbit. If it is further assumed that motion in the transfer orbit does not deviate significantly from these neighboring orbits, linearization of the equations of motion is permissible.

The analysis has been carried out in three sets of variables: (1) rotating rectangular coordinates, (2) rotating spherical coordinates, and (3) Lagrange's planetary variables. The rotating coordinates have been utilized previously in Refs. 5, 6, and 7, while the planetary variables were applied to an orbit transfer problem in Ref. 4.

The rotating coordinate systems are depicted in Figs. 1 and 2. Each consists of an origin which revolves at satellite velocity in the initial (interior) circular orbit and orthogonal coordinates measured from this revolving origin. In the rectangular system of Fig. 1,  $y'$  is a radial dimension,  $x'$  is measured tangent to the initial orbit at the origin, and  $z'$  is a coordinate which is out of the plane of the initial orbit and is normal to both  $x'$  and  $y'$ .

In Fig. 2, the spherical system is composed of a radial coordinate  $y$ , an arc  $x$ , measured circumferentially from the origin, and another arc  $z$ , which is orthogonal to the  $x$ - $y$  plane.

The Lagrange planetary variables, which are derived from the elements of an elliptic orbit and are used in the standard variation-of-parameters equations of celestial mechanics (Ref. 8), are convenient because they eliminate the necessity of treating singularities for zero eccentricity and zero inclination in these equations. As they are used in this study, the planetary variables consist of the nondimensionalized semi-major axis  $x_1 = a/a_0$ , a circumferential distance component,  $x_4$ , and the following combinations of the remaining orbital elements:

$$\begin{aligned} x_2 &= e \sin \omega \\ x_3 &= e \cos \omega \\ x_5 &= \sin i \sin \Omega \\ x_6 &= \sin i \cos \Omega \end{aligned} \tag{1}$$

in that "fast" trajectories are allowed only when the linearizing assumptions may be relaxed. On the other hand, fast trajectories are allowed in the rectangular system because no limits are placed on the component velocities in the linearizing process.

### Analysis

The optimization problem is to derive the optimal control equation for the minimum-fuel transfer or rendezvous of a power-limited rocket between neighboring orbits in a given time. Mathematically, this requires minimization of the integral

$$J = \int_0^{t_f} (T/m)^2 dt = \int_0^{\tau_f} (n_0/2) A^2 d\tau = \int_0^{\tau_f} f_0(A) d\tau \quad (2)$$

subject to constraints imposed by the equations of state which may be expressed in the form

$$\dot{x}_i = f_i(x, A) \quad i = 1, \dots, n \quad (3)$$

The control is the thrust acceleration vector,  $A$ , in the present case.

The problem is treated as a problem of Lagrange in the calculus of variations. In particular, Breakwell's formulation (Ref. 9) of this problem is used because the linearized equations in the present case are particularly well suited to this formulation.

If a fundamental function  $F$  is defined as

$$F = -f_0 + \sum_{i=1}^n \lambda_i f_i \quad (4)$$

the variational treatment requires satisfaction of Euler-Lagrange equations in the following form as necessary conditions for the existence of an extremal arc:

$$\frac{d\lambda_i}{d\tau} = - \frac{\partial F}{\partial x_i} \quad (5)$$

$$\frac{\partial F}{\partial A_j} = 0 \quad (6)$$

where  $e$  is eccentricity,  $w$  is the longitude of peri-apsis,  $i$  is orbital inclination, and  $\Omega$  is the longitude of the ascending node. The planetary variables provide a simple means of introducing eccentricity into the terminal orbits, and the form of the state equations using these variables is particularly simple in the present problem. However, in a practical application, they might be less desirable than the rotating coordinates because the orbital elements cannot be directly measured.

In view of the foregoing considerations, eccentric terminal orbits have been allowed only in the planetary variables in this study, while the analysis in rotating reference frames is confined to circular terminal orbits.

It should be noted here that the three sets of variables are entirely equivalent in that the equations of motion may be transformed directly from one set to another by substitution. There are some differences in the required linearizing assumptions which should be mentioned, however.

Consider the coordinate system depicted in Fig. 1, a rectangular system with its origin fixed on the interior orbit (assumed to be the reference orbit) in the  $x', y'$  plane. The mutually orthogonal coordinates  $x', y'$ , and  $z'$  form a triad that revolves with angular speed  $n_0$  characteristic of the reference orbit, so that motion in this frame of reference is relative to a point on the reference orbit. The spherical coordinate system in Fig. 2 is described by the arc  $x$  in the plane of the reference orbit, the arc  $z$  measured normal to this plane, and a radial dimension  $y$ .

In order to linearize the equations of motion in the first system, it is necessary to assume that excursions  $x', y'$ , and  $z'$  from the origin be small in comparison with the radius,  $r_0$ , of the reference orbit. Motion is therefore constrained to a small sphere about the origin. No restrictions are placed on the component velocities. In the rotating spherical system, only the assumption of small component velocities will linearize the equations, whereas the arc  $x$  is not limited. The resultant motion is constrained to a torus about the reference orbit.

Since the linearized equations of motion are identical except for differences in notation (Ref. 5), one can draw the conclusion that, if in the spherical system the resultant motion does not involve large variations in  $x$ , the velocity components may be large. In the present study, use of the spherical system has been assumed throughout, and the results may be extended according to the foregoing discussion.

In the case of the planetary variables, the linearizing assumptions require that the difference in the semi-major axes of the terminal orbits be small and that the eccentricity of the terminal orbits as well as the eccentricity of the instantaneous transfer orbit be small. The implications of these assumptions are similar to those for the rotating spherical system

An additional necessary condition provided by the Pontryagin Maximum Principle must also be satisfied to ensure that the stationary solution predicted by the Euler equations is actually an extremum. The maximum principle, which may be expressed as

$$F(x_i, \lambda_i, A_j^*) \geq F(x_i, \lambda_i, A_j) \quad (7)$$

ensures that the stationary solution is an absolute maximum. Furthermore, it has been shown (Ref. 10) that for a system where both the state variables and the controls appear linearly in the state equations, the maximum principle is also sufficient to ensure a minimum of the payoff,  $J$ . Since all cases in the present analyses are linear in the controls and satisfy the maximum principle, the optimum trajectories described herein are absolute extrema.

Due to the great number of equations involved, the variational analysis is not described in each case. Only the most important equations are included, and these are grouped in an orderly fashion in the appendixes. The rotating coordinate systems are considered in Appendix I, and the planetary variables are considered in Appendix II. For a more detailed account of the application of the aforementioned equations the reader is referred to Ref. 2 wherein a specific case is treated in detail.

### Synthesis of the Optimal Controls

In order to put the equations for the optimized controls into a form compatible with guidance requirements, several changes are made. First,  $\tau$  in the control equations is replaced by  $-\tau$ . That is, the equations are rewritten with "time-to-go" as the independent variable. Secondly, while in the ordinary transfer and rendezvous analyses in rotating coordinates it was generally convenient to assume zero initial conditions, the terminals are reversed in the control synthesis. That is, the target orbit is assumed to be defined by zero values in most of the state variables. The results of the control synthesis are expressed in terms of the guidance coefficients,  $\partial A_j / \partial x_i$ , of each component of the control vector,  $A$ .

The equations for the control synthesis are summarized in Appendix III for transfer and rendezvous in each of the coordinate systems. Those equations which deal specifically with transfer between circular orbits have been presented previously in Ref. 3.

## RESULTS

## Orbit Transfer and Rendezvous

The multiplicity of solutions generated in this study (particularly for rendezvous) precludes a graphical presentation of all the resulting trajectories. An attempt is made to summarize the results in a reasonably concise form with orbit transfer solutions represented as special cases of rendezvous wherever feasible.

To simplify the presentation of the results, only circle-to-circle transfer and rendezvous cases are examined in the summary curves of Figs. 3 through 13. The first set of plots, Figs. 3 through 5, shows the variation of the components of the optimal thrust acceleration with time for circle-to-circle transfer only.

The in-plane components  $A_x/y_f$  and  $A_y/y_f$  are seen to display symmetry about the midpoint in time for all trip times, as does the out-of-plane component  $A_z/r_{0i}$ . In particular, when  $\tau_f = 2n\pi$ , the components  $A_x/y_f$  and  $A_y/y_f$  are constant with time, and the latter is zero. For the coplanar problem, constant circumferential thrust acceleration is thereby specified as the optimum mode for integral multiples of the period of the reference orbit, a result that is in agreement with Ref. 7.

Figures 6 through 8 show the thrust acceleration components for circle-to-circle rendezvous at a particular trip time equal to one sixth of an orbital period of the reference orbit. The parameter in Figs. 6 and 7 is  $x_f/y_f\tau_f$  which takes on the value of  $3/4$  for the special case of optimum transfer. Similarly the out-of-plane component is plotted with  $\Omega_f$  as a parameter. As indicated, the longitude of the node can have either of two values, 150 or 330 deg, for optimum transfer.

The payoff,  $J$ , can be best represented as the sum of three components,  $J_1$ ,  $J_2$ , and  $J_3$ , which are defined by Eqs. (A-44) and (A-45) and are plotted in Figs. 9 through 11. The components  $J_1$  and  $J_2$  define propellant requirements for coplanar rendezvous, while the addition of  $J_3$  introduces the out-of-plane requirement. In particular  $J$  is equal to  $J_1$  for coplanar transfer since the term  $x_f/y_f\tau_f - 3/4$  in  $J_2$  is zero for optimum transfer.

All three components, as well as their sum, are seen to be monotonically decreasing functions of  $\tau_f$ . In the limit, as  $\tau_f \rightarrow \infty$ ,  $A$  and  $J \rightarrow 0$ . This is a consequence of the fact that no limit has been placed on exhaust velocity. Similarly all three components tend to infinity as  $\tau_f$  approaches zero because zero trip time requires infinite thrust acceleration.

An interesting feature of  $J_3$  is evident from Fig. 11. For  $\tau_f = k\pi$ , where  $k = 0, 1, 2, \dots$ ,  $J_3$  is the same for all nodal longitudes,  $\Omega_f$ . For all other

times the envelope of the family of curves is given by the equations

$$J_{3\max} = \frac{1}{\tau_f - |\sin \tau_f|} \quad (8)$$

$$J_{3\min} = \frac{1}{\tau_f + |\sin \tau_f|} \quad (9)$$

where the lower envelope is given by Eq. (9) and represents  $J_3$  for optimum transfer.

#### Choice of Reference Orbit

It has been observed that the linearized equations are applicable only for orbits which are not separated by large radial distances. More specifically, excursions from the origin in the  $y$  direction should always be small. It is apparent, however, that when the reference orbit is chosen to have the same radius as the initial orbit the excursion,  $y$ , to the final orbit is maximized. A better reference orbit would be one midway between the terminal orbits since this device would guarantee a radial excursion no greater than half the distance between the terminals.

Although for the most part, the equations of this report are based on a reference orbit coincident with the initial orbit, Eqs. (A-48) through (A-51) and (A-131) through (A-134) are exceptions in this respect. These equations are derived to account for an arbitrary choice of the reference orbit and may therefore be applicable in cases where the ordinary equations break down.

#### Application to Planetary Orbits

Strictly speaking, none of the planetary orbits are "neighboring orbits" in the sense in which this term has been defined. Earth's closest neighbor, Venus, has a semi-major axis,  $a = 0.7233\text{AU}$  compared with  $a = 1.0\text{AU}$  for earth, leaving a separation distance of  $0.2767\text{AU}$  which is not  $\ll 1.0\text{AU}$ . However, using the improvement referred to above, it is possible to apply the linearized analysis to earth-Venus and earth-Mars trajectories with remarkably good accuracy. In Figs. 12 and 13, comparisons have been made with exact solutions from Ref. 11, for earth-Venus and earth-Mars transfers. The circled points were calculated from Eq. (A-48) of Appendix I using a reference orbit midway between the two terminal orbits. These results for the special case of uninclined, circular terminal orbits show only a slight discrepancy in  $J$  for transfer times up to one earth year.

### Extension of the Linearized Theory

Based on the successful correlation indicated by Figs. 12 and 13, a new theory is being considered in order to account for second-order effects in  $J$ . This theory is a "piecewise-linear" analysis which may be described as follows: The transfer (or rendezvous) is divided into two steps, each requiring a portion of the total trip time. The first segment of the trajectory consists of a rendezvous from the initial orbit to an intermediate orbit of unspecified size and shape, and the second segment is a rendezvous from this intermediate orbit to the final terminal orbit. The expression for  $J$  is composed of two linear expressions for the two segments, and the parameters of the intermediate orbit are considered as variables which may be optimized so as to minimize the total  $J$ . In each segment an appropriate reference orbit is chosen so as to improve the accuracy of the theory.

This approach should provide better results than the linearized theory. Since the results for earth-Mars and earth-Venus transfers were already good, the piecewise-linear theory may approach exact results in these cases and might even yield acceptable results for trajectories to the outer planets.

### Control Synthesis

In this study it has been possible to express each of the components of the optimal control vector,  $A$ , as a linear function of the  $n$  state variables.

$$A_j = \sum_{i=1}^n \frac{\partial A_j}{\partial x_i} x_i \quad (10)$$

Therefore, the presentation of the results can be confined to curves of the guidance coefficients,  $\partial A_j / \partial x_i$  plotted against time to go,  $\tau'$ . Using the equations for the guidance coefficients which comprise Appendix III, the summary curves of Figs. 14 through 25 were generated.

The synthesized controls for the case of transfer between an arbitrary state and a nearby circular orbit appear in Figs. 14 through 16 in terms of the rotating coordinate system variables. The extension to include eccentricity of the final orbit is provided by use of the Lagrange planetary variables in Figs. 17 through 19.

For rendezvous the same procedure is followed in the presentation of the synthesized controls, with the addition of curves to account for the dependence of in-plane thrust acceleration components on the circumferential distance. In rotating coordinates, Figs. 20 through 22 summarize the results for rendezvous between any initial state and a point on a nearby circular orbit.

As in the transfer case, the planetary variables facilitate the extension to rendezvous between an initial state and a point on a nearby orbit of low eccentricity. The results for the planetary variables appear in Figs. 23 through 25.

All the curves for the guidance coefficients display similar behavior. When time-to-go is short, the curves diverge to infinity (either positive or negative), but a damped oscillation is evident, causing the coefficients to approach zero for very long times.



## REFERENCES

1. Gobetz, F. W.: Optimal Variable-Thrust Rendezvous of a Power-Limited Rocket Between Neighboring Low-Eccentricity Orbits (Interim Report), Progress Report No. 6, Studies in the Fields of Space Flight and Guidance Theory, NASA, MSFC (in preparation).
2. Gobetz, F. W.: Optimal Variable-Thrust Transfer of a Power-Limited Rocket Between Neighboring Circular Orbits. AA Journal, Vol. 2, No. 2, pp. 339-343, February 1964.
3. Gobetz, F. W.: Control Synthesis for Optimal Low-Thrust Transfer Between a Circular Orbit and a Nearby Orbit. Presented at the Fall 1963 Meeting of SIAM, New York.
4. Carlson, N. A.: Optimal Constant-Thrust Transfer Between Adjacent Coplanar Circular Orbits. UAC Research Laboratories Report B-110058-12, November 1963.
5. Kelley, H. J. and J. C. Dunn: An Optimal Guidance Approximation for Quasi-Circular Orbital Rendezvous. Second IFAC Congress on Automatic Control, Basle, Switzerland, September 1963.
6. Clohessy, W. H. and R. S. Wiltshire: Terminal Guidance System for Satellite Rendezvous. Journal of Aerospace Sciences, Vol. 27, pp. 653-658, 1960.
7. Hinz, H. K.: Optimal Low-Thrust Near-Circular Orbit Transfer. AIAA Journal, Vol. 1, No. 6, pp. 1367-1371, June 1963.
8. Moulton, F. R.: An Introduction to Celestial Mechanics. The MacMillan Co., New York, 1914.
9. Breakwell, J. V.: The Optimization of Trajectories. SIAM Journal, Vol. 7, pp. 215-222, June 1959.
10. Rozonoer, L. I., and L. S. Pontryagin: Maximum Principle in the Theory of Optimum Systems, Part I, Automation and Control, Vol. 20, No. 10, October 1959.
11. Melbourne, W. G.: Interplanetary Trajectories and Payload Capabilities of Advanced Propulsion Vehicles. JPL TR 32-68, March 1961.

## LIST OF SYMBOLS

$\frac{T}{m}$	Thrust-to-mass ratio
$A$	$\frac{1}{n_o} \frac{T}{m}$
$C$	Integration constant
$f$	Rate of change of a state variable
$F$	Fundamental function
$J$	Defined by Eq. (2)
$D$	Defined by Eq. (A-154)
$B$	Defined by Eq. (A-182)
$Q$	Defined by Eq. (A-181)
$\Phi$	Defined by Eq. (A-146)
$\lambda$	Lagrange multiplier
$r$	Radius
$R$	Radial force
$W$	Normal force
$S$	Circumferential force
$n$	Mean angular motion
$x, y, z$	Position components in spherical system
$x', y', z'$	Position components in rectangular system
$u, v, w$	Velocity components in $x, y, z$ , directions
$t$	Time
$\tau$	$n_o t$

LIST OF SYMBOLS  
(contd.)

$\tau'$	Time to go
$\eta$	True anomaly
$\omega$	Longitude of peri-apsis
$e$	Eccentricity
$N$	Unit vector normal to instantaneous transfer orbit
$a$	Semi-major axis
$\Omega$	Longitude of the node
$i$	Inclination
$x_1$	$a/a_0$
$x_2$	$e \sin \omega$
$x_3$	$e \cos \omega$
$x_5$	$\sin i \sin \Omega$
$x_6$	$\sin i \cos \Omega$
$\vec{c}$	Angular momentum vector

Subscripts

$i$	Index denoting $x, y, z, u, v, w$
$j$	Index denoting $x, y, z$
$o$	Initial condition
$f$	Final condition
$x, y, z, u, v, w$	Denoting state variable
$I$	Intermediate reference orbit
$R$	Radial

LIST OF SYMBOLS  
(contd.)

S                    Circumferential

W                    Normal

Superscripts

\*                    Optimum condition

→                    Denotes a vector

## APPENDIX I

## ROTATING RECTANGULAR AND SPHERICAL COORDINATE SYSTEMS

1. Equations of State

$$\frac{dx}{d\tau} = u \quad (A-1)$$

$$\frac{dy}{d\tau} = v \quad (A-2)$$

$$\frac{dz}{d\tau} = w \quad (A-3)$$

$$\frac{du}{d\tau} = A_x + 2y \quad (A-4)$$

$$\frac{dv}{d\tau} = A_y + 3y - 2u \quad (A-5)$$

$$\frac{dw}{d\tau} = A_z - z \quad (A-6)$$

2. Euler-Lagrange Equations

$$\dot{\lambda}_x = 0 \quad (A-7)$$

$$\dot{\lambda}_y = -3\lambda_v \quad (A-8)$$

$$\dot{\lambda}_z = \lambda_w \quad (A-9)$$

$$\dot{\lambda}_u = -\lambda_x + 2\lambda_v \quad (A-10)$$

$$\dot{\lambda}_v = -\lambda_y - 2\lambda_u \quad (A-11)$$

$$\dot{\lambda}_w = -\lambda_z \quad (A-12)$$

$$\lambda_u = n_0 A_x \quad (A-13)$$

$$\lambda_v = n_0 A_y \quad (A-14)$$

$$\lambda_w = n_0 A_z \quad (A-15)$$

### 3. Integrated Euler-Lagrange Equations

$$\lambda_x = n_0 C_0 \quad (A-16)$$

$$\lambda_y = -6n_0(C_4 + C_0\tau - C_1 \cos\tau + C_2 \sin\tau) \quad (A-17)$$

$$\lambda_z = 2n_0(C_3 \sin\tau + C_3 \cos\tau) \quad (A-18)$$

$$\lambda_u = n_0(3C_4 + 3C_0\tau - 4C_1 \cos\tau + 4C_2 \sin\tau) \quad (A-19)$$

$$\lambda_v = 2n_0(C_0 + C_1 \sin\tau + C_2 \cos\tau) \quad (A-20)$$

$$\lambda_w = 2n_0(C_3 \cos\tau - C_3 \sin\tau) \quad (A-21)$$

### 4. Boundary Conditions

<u>State Variable</u>	<u>Transfer</u>		<u>Rendezvous</u>	
	$\tau = 0$	$\tau = \tau_f$	$\tau = 0$	$\tau = \tau_f$
x	0	FREE	0	$x_f$
y	0	$y_f$	0	$y_f$
z	0	$z_f$	0	$z_f$
u	0	$\frac{3}{2} y_f^{(1)}$	0	$\frac{3}{2} y_f^{(1)}$
v	0	0	0	0
w	0	$\sqrt{r_0^2 - z_f^2}^{(2)}$	0	$\sqrt{r_0^2 - z_f^2}^{(2)}$

### 5. Integrated Equations of State (with initial conditions)

$$\begin{aligned} x = & \left[ 16(\tau - \sin\tau) - \frac{3}{2}\tau^3 \right] C_0 + \left[ 16(1 - \cos\tau) - 10\tau \sin\tau \right] C_1 \\ & + \left[ 22 \sin\tau - 10\tau \cos\tau - 12\tau \right] C_2 - \left[ \frac{9}{2}\tau^2 - 12(1 - \cos\tau) \right] C_4 \end{aligned} \quad (A-22)$$

$$\begin{aligned} y = & \left[ 8(1 - \cos\tau) - 3\tau^2 \right] C_0 + 5 \left[ \sin\tau - \tau \cos\tau \right] C_1 + \left[ 5\tau \sin\tau - 8(1 - \cos\tau) \right] C_2 \\ & + 6 \left[ \sin\tau - \tau \right] C_4 \end{aligned} \quad (A-23)$$

(1) REF 6

(2) REF 5

$$z = [\tau \cos \tau - \sin \tau] C_3 + [\tau \sin \tau] C_5 \quad (A-24)$$

$$u = \left[ 16(1 - \cos \tau) - \frac{9}{2} \tau^2 \right] C_0 + [6 \sin \tau - 10 \tau \cos \tau] C_1 \\ + [10 \tau \sin \tau - 12(1 - \cos \tau)] C_2 + [12 \sin \tau - 9 \tau] C_4 \quad (A-25)$$

$$v = [8 \sin \tau - 6 \tau] C_0 + [5 \tau \sin \tau] C_1 + [5 \tau \cos \tau - 3 \sin \tau] C_2 \\ + 3 [1 - \cos \tau] C_4 \quad (A-26)$$

$$w = [-\tau \sin \tau] C_3 + [\sin \tau + \tau \cos \tau] C_5 \quad (A-27)$$

#### 6. Transversality Conditions - Transfer

$$\lambda_x = C_0 = 0 \quad (A-28)$$

$$\frac{C_5}{C_3} = \frac{\tan \tau_f + \frac{w_f}{z_f}}{1 - \frac{w_f}{z_f} \tan \tau_f} \quad (A-29)$$

#### 7. Constants of Integration - Transfer

$$C_1 = \frac{y_f \sin \tau_f}{16(1 - \cos \tau_f) - \tau_f(5\tau_f + 3 \sin \tau_f)} \quad (A-30)$$

$$C_2 = \frac{-y_f(1 - \cos \tau_f)}{16(1 - \cos \tau_f) - \tau_f(5\tau_f + 3 \sin \tau_f)} \quad (A-31)$$

$$C_3 = \frac{(\sin \tau_f + \tau_f \cos \tau_f) z_f - (\tau_f \sin \tau_f) \sqrt{r_0^2 i^2 - z_f^2}}{\tau_f^2 - \sin^2 \tau_f} \quad (A-32)$$

$$C_4 = \frac{\frac{y_f}{6} (5\tau_f + 3\sin\tau_f)}{16(1 - \cos\tau_f) - \tau_f(5\tau_f + 3\sin\tau_f)} \quad (A-33)$$

Rendezvous

$$C_0 = \frac{\tau_f y_f \left( \frac{x_f}{y_f \tau_f} - \frac{3}{4} \right) (5\tau_f - 3\sin\tau_f)}{\frac{3}{4} \tau_f (5\tau_f - 3\sin\tau_f)(\tau_f^2 - 80) + 4(1 - \cos\tau_f)(71\tau_f^2 - 64) + 248\tau_f^2 \cos\tau_f} \quad (A-34)$$

$$C_1 = \frac{y_f \sin\tau_f}{16(1 - \cos\tau_f) - \tau_f(5\tau_f + 3\sin\tau_f)} + C_0 \left[ \frac{3\sin\tau_f - 8(1 - \cos\tau_f)}{5\tau_f - 3\sin\tau_f} \right] \quad (A-35)$$

$$C_2 = \frac{-y_f(1 - \cos\tau_f)}{16(1 - \cos\tau_f) - \tau_f(5\tau_f + 3\sin\tau_f)} + C_0 \left[ \frac{3\tau_f(1 + \cos\tau_f) - 8\sin\tau_f}{5\tau_f - 3\sin\tau_f} \right] \quad (A-36)$$

$$C_3 = \frac{(\sin\tau_f + \tau_f \cos\tau_f) z_f - (\tau_f \sin\tau_f) \sqrt{r_0^2 i^2 - z_f^2}}{(\tau_f^2 - \sin^2\tau_f)} \quad (A-37)$$

$$C_4 = \frac{\frac{y_f}{6} (5\tau_f + 3\sin\tau_f)}{16(1 - \cos\tau_f) - \tau_f(5\tau_f + 3\sin\tau_f)} - C_0 \frac{\tau_f}{2} \quad (A-38)$$

$$C_5 = \frac{(\tau_f \sin\tau_f) z_f + (\tau_f \cos\tau_f - \sin\tau_f) \sqrt{r_0^2 i^2 - z_f^2}}{(\tau_f^2 - \sin^2\tau_f)} \quad (A-39)$$

## 8. Controls

$$A_x = 3C_4 + 3C_0\tau - 4C_1 \cos\tau + 4C_2 \sin\tau \quad (A-40)$$



$$A_y = 2 \left[ C_0 + C_1 \sin \tau + C_2 \cos \tau \right] \quad (A-41)$$

$$A_z = 2 \left[ C_5 \cos \tau - C_3 \sin \tau \right] \quad (A-42)$$

## 9. Payoff

### Transfer

$$\frac{J}{n_o^3 r_o^2} = \frac{\left( \frac{y_f}{r_o} \right)^2 (5\tau_f + 3\sin \tau_f)}{8 \left[ \tau_f (5\tau_f + 3\sin \tau_f) - 16(1 - \cos \tau_f) \right]} + \frac{i^2}{\tau_f + |\sin \tau_f|} \quad (A-43)$$

### Rendezvous

$$\frac{J}{n_o^3 r_o^2} = J_1 \left( \frac{y_f}{r_o} \right)^2 + J_2 \left( \frac{y_f}{r_o} \right)^2 \left( \frac{x_f}{y_f \tau_f} - \frac{3}{4} \right)^2 + J_3 i^2 \quad (A-44)$$

$$\begin{aligned} \frac{J}{n_o^3 r_o^2} = & \frac{\left( \frac{y_f}{r_o} \right)^2 (5\tau_f + 3\sin \tau_f)}{8 \left[ \tau_f (5\tau_f + 3\sin \tau_f) - 16(1 - \cos \tau_f) \right]} \\ & + \frac{\frac{\tau_f^2}{2} \left( \frac{y_f}{r_o} \right)^2 \left( \frac{x_f}{y_f \tau_f} - \frac{3}{4} \right)^2 (5\tau_f - 3\sin \tau_f)}{\frac{3}{4} \tau_f (5\tau_f - 3\sin \tau_f) (\tau_f^2 - 80) + 4(1 - \cos \tau_f) (71\tau_f^2 - 64) + 248\tau_f^2 \cos \tau_f} \\ & + i^2 \left[ \frac{\tau_f - \sin \tau_f \cos (2\Omega_f + \tau_f)}{(\tau_f^2 - \sin^2 \tau_f)} \right] \end{aligned} \quad (A-45)$$

10. It should be pointed out that for each free end condition in the case of orbit transfer, the variational analysis predicts an optimum value for that particular state variable at the end point. In the rotating coordinate systems the x and z coordinates are left open at final time,  $\tau_f$ . The end point for the optimal transfer is then determined in the analysis and is defined by the equations.

$$\left(\frac{z_f}{r_o}\right)^* = i \sqrt{\frac{1 \mp \cos \tau_f}{2}} \quad (A-46)$$

$$\left(\frac{x_f}{y_f}\right)^* = \frac{3}{4} \tau_f \quad (A-47)$$

### 11. Payoff Equations with an Intermediate Reference Orbit

Let the origin revolve in a circular orbit of radius  $r_I$  between the two terminal orbits such that the radial distance to the outer orbit is  $r_f - r_I$  and the radial distance to the inner orbit is  $r_I - r_o$ . The radii  $r_o$  and  $r_f$  refer to the inner and outer orbits, respectively.

#### Transfer

$$\frac{J}{n_I^3 r_I^2} = \frac{\frac{1}{8} \left(\frac{r_f - r_o}{r_I}\right)^2 (5\tau_f + 3\sin\tau_f)}{\tau_f (5\tau_f + 3\sin\tau_f) - 16(1 - \cos\tau_f)} + \frac{i^2}{\tau_f + |\sin\tau_f|} \quad (A-48)$$

#### Rendezvous

$$\begin{aligned} \frac{J}{n_I^3 r_I^2} &= \frac{\frac{1}{8} \left(\frac{r_f - r_o}{r_I}\right)^2 (5\tau_f + 3\sin\tau_f)}{\tau_f (5\tau_f + 3\sin\tau_f) - 16(1 - \cos\tau_f)} \\ &+ \frac{\frac{\tau_f^2}{2} \left\{ \frac{x_f}{\tau_f r_I} - \frac{3}{4} \left( \frac{r_f + r_o}{r_I} - 2 \right) \right\}^2 (5\tau_f - 3\sin\tau_f)}{\frac{3}{4} \tau_f (5\tau_f - 3\sin\tau_f) (\tau_f^2 - 80) + 4(1 - \cos\tau_f) (7\tau_f^2 - 64) + 248\tau_f^2 \cos\tau_f} \\ &+ i^2 \left\{ \frac{\tau_f^2 - \sin\tau_f \cos(2\Omega_f + \tau_f)}{\tau_f^2 - \sin^2\tau_f} \right\} \end{aligned} \quad (A-49)$$

12. Optimal Transfer Coordinates

$$\left(\frac{x_f}{r_I}\right)^* = \frac{3}{4} \tau_f \left(\frac{r_f + r_o}{r_I} - 2\right) \quad (A-50)$$

$$\left(\frac{z_f}{r_I}\right)^* = i \sqrt{\frac{1 \pm \cos \tau_f}{2}} \quad (A-51)$$

## APPENDIX II

## LAGRANGE'S VARIABLES

In the theory of special perturbations, as derived in Ref. 8 for example, the equations for rates of change of the elements of an elliptic orbit are written in terms of the elements and acceleration components S, R, and W, which are perpendicular to the radius vector, radial and normal to the orbital plane, respectively.

Consider the five elements, a, e, i,  $\omega$ ,  $\Omega$ . The equations for small rates of change of these variables are

$$\frac{da}{dt} = \frac{2}{n\sqrt{1-e^2}} \left[ eR \sin \eta + S(1 + e \cos \eta) \right] \quad (A-52)$$

$$\frac{de}{dt} = \frac{\sqrt{1-e^2}}{na} \left[ R \sin \eta + \frac{2 \cos \eta + e + e \cos^2 \eta}{1 + e \cos \eta} S \right] \quad (A-53)$$

$$\frac{di}{dt} = \frac{\sqrt{1-e^2}}{na} W \cos(\omega + \eta) \quad (A-54)$$

$$\frac{d\omega}{dt} = \frac{\sqrt{1-e^2}}{nae} \left[ -R \cos \eta + \frac{2 + e \cos \eta}{1 + e \cos \eta} S \sin \eta - \frac{e \tan \frac{i}{2} \sin(\omega + \eta)}{1 + e \cos \eta} W \right] \quad (A-55)$$

$$\frac{d\Omega}{dt} = \frac{\sqrt{1-e^2}}{na} \frac{W}{\sin i} \sin(\omega + \eta) \quad (A-56)$$

In order to avoid singularities for zero eccentricity and inclination in Eqs. (A-55) and (A-56) these equations may be transformed according to the following definitions:

$$x_2 = e \sin \omega \quad (A-57)$$

$$x_3 = e \cos \omega \quad (A-58)$$

$$x_5 = \sin i \sin \Omega \quad (A-59)$$

$$x_6 = \sin i \cos \Omega \quad (A-60)$$

Under the assumptions

$$\begin{aligned} e &\ll 1 \\ a &\approx a_0 \\ n &\approx n_0 \\ \tau &= n_0 t = \omega + \eta \\ i &\ll 1 \end{aligned} \quad (A-61)$$

$$A_R = \frac{R}{a_0 n_0^2}, \quad A_S = \frac{S}{a_0 n_0^2}, \quad A_W = \frac{W}{a_0 n_0^2} \quad (A-62)$$

and with the further definitions

$$x_1 = \frac{a}{a_0} \quad (A-63)$$

$$x_4 = x \quad (A-64)$$

the equations of state for the variational problem may be derived from Eqs. (A-52) through (A-60).

There is a direct equivalence between these equations and the equations of state in the rotating coordinate system variables. That is, each of the Lagrange variables  $x_1, x_2, x_3, \dots, x_6$ , can be expressed in terms of the rotating coordinate variables,  $x, y, z, u, v$ , and  $w$ .

Referring to Fig. 26, define a position vector  $\vec{r}$  in nonrotating coordinates originating at the center of attraction  $F$ . Assume the motion out of the reference plane is uncoupled from the in-plane motion.

Relative to a rotating rectangular coordinate system originating at  $O$  and rotating with angular velocity  $\vec{n}$  this vector is

$$\vec{r} = x\vec{i} + (r_0 + y)\vec{j} \quad (\text{A-65})$$

where the unit vectors  $\vec{i}$  and  $\vec{j}$  are taken in the x and y directions, respectively. The vector velocity  $\vec{V}$  is obtained by differentiating  $\vec{r}$ .

$$\vec{V} = \frac{d\vec{r}}{dt} = u\vec{i} + v\vec{j} + \dot{n}x\vec{i} \quad (\text{A-66})$$

Since  $\vec{n} = n_0\vec{k}$ , the expression for  $\vec{V}$  is

$$\vec{V} = [u - n_0(r_0 + y)]\vec{i} + (v + n_0x)\vec{j} \quad (\text{A-67})$$

Using Eqs. (A-65) and (A-67), expressions can be written for the angular momentum  $\vec{C}$ , the path speed  $V$  and the radius  $r$  of the vehicle

$$\vec{C} = \vec{r} \times \vec{V} = [x(v + n_0x) - (r_0 + y)(u - n_0(r_0 + y))]\vec{k} \quad (\text{A-68})$$

$$V = \sqrt{\vec{V} \cdot \vec{V}} = \sqrt{[u - n_0(r_0 + y)]^2 + [v + n_0x]^2} \quad (\text{A-69})$$

$$r = \sqrt{\vec{r} \cdot \vec{r}} = \sqrt{x^2 + (r_0 + y)^2} \quad (\text{A-70})$$

The following equations can be written for the angular momentum, speed, and radius of a body in an inverse square field.

$$|\vec{C}| = \sqrt{Ka(1 - e^2)} \quad (\text{A-71})$$

$$V = \sqrt{K\left(\frac{2}{r} - \frac{1}{a}\right)} = \sqrt{(\dot{r})^2 + (r\dot{\eta})^2} \quad (\text{A-72})$$

$$r = \frac{a(1 - e^2)}{1 + e \cos \eta} \quad (\text{A-73})$$

Combining these equations with the absolute value of  $\vec{C}$ , and with  $V$  and  $r$  from Eqs. (A-68), (A-69), and (A-70), the following scalar equations result.

$$\frac{a}{a_0} = \left(1 + \frac{y}{r_0}\right)(1 + e \cos \eta) \quad (\text{A-74})$$

$$\frac{u}{n_0 r_0} - \left(1 + \frac{y}{r_0}\right) = \frac{\sqrt{\frac{a}{a_0}}}{1 + \frac{y}{r_0}} \quad (\text{A-75})$$

$$\frac{v}{n_0 r_0} + \frac{x}{r_0} = \sqrt{\frac{e \cos \eta}{\frac{a}{a_0}}} \quad (\text{A-76})$$

Finally, noting that

$$\begin{aligned} \frac{a}{a_0} &= x_1, \quad x_2 = e \sin \omega, \quad x_3 = e \cos \omega \\ e \cos \eta &= e \cos(\tau - \omega) = x_2 \sin \tau + x_3 \cos \tau \end{aligned} \quad (\text{A-77})$$

the equations relating the coordinates are obtained.

$$\frac{y}{r_0} = (x_1 - 1) - x_2 \sin \tau - x_3 \cos \tau \quad (\text{A-78})$$

$$\frac{v}{n_0 r_0} = x_3 \cos \tau - x_2 \sin \tau \quad (\text{A-79})$$

$$\frac{u}{n_0 r_0} = \frac{3}{2} (x_1 - 1) - 2x_2 \sin \tau - 2x_3 \cos \tau \quad (\text{A-80})$$

The components of the out-of-plane motion can be related in the following way. If  $\vec{N}$  is a unit vector normal to the instantaneous transfer orbit and  $\vec{s}$  is a unit vector in the direction of the line of nodes, then

$$\vec{s} = \vec{N} \times \vec{k} \quad (\text{A-81})$$

and, since the angle between  $\vec{s}$  and the vehicle is  $\tau - \Omega$ ,

$$\cos(\tau - \Omega) = \vec{s} \cdot \vec{i} \quad (\text{A-82})$$

Also, the orbital inclination is

$$\cos i = \vec{N} \cdot \vec{k} \quad (\text{A-83})$$

Using these parameters the equation for the elevation,  $z$ , of the probe is

$$\frac{z}{r_0} = \tan i \sin(\tau - \Omega) \approx \sin i \sin(\tau - \Omega) \quad (\text{A-84})$$

or

$$\frac{z}{r_0} = -x_5 \cos \tau + x_6 \sin \tau \quad (\text{A-85})$$

The out-of-plane velocity,  $w$ , is

$$\frac{w}{r_0 \dot{\tau}_0} = x_5 \sin \tau + x_6 \cos \tau \quad (\text{A-86})$$

### 1. Equations of State

$$\frac{dx_1}{d\tau} = 2A_S \quad (\text{A-87})$$

$$\frac{dx_2}{d\tau} = 2A_S \sin \tau - A_R \cos \tau \quad (\text{A-88})$$

$$\frac{dx_3}{d\tau} = 2A_S \cos \tau + A_R \sin \tau \quad (\text{A-89})$$

$$\frac{dx_4}{d\tau} = \frac{3}{2}(x_1 - 1) - 2x_2 \sin \tau - 2x_3 \cos \tau \quad (\text{A-90})$$

$$\frac{dx_5}{d\tau} = -A_W \sin \tau \quad (\text{A-91})$$

$$\frac{dx_6}{d\tau} = A_W \cos \tau \quad (\text{A-92})$$



2. Euler-Lagrange Equations

$$\dot{\lambda}_1 = -\frac{3}{2} \lambda_4 \quad (\text{A-93})$$

$$\dot{\lambda}_2 = 2\lambda_4 \sin \tau \quad (\text{A-94})$$

$$\dot{\lambda}_3 = 2\lambda_4 \cos \tau \quad (\text{A-95})$$

$$\dot{\lambda}_4 = \dot{\lambda}_5 = \dot{\lambda}_6 = 0 \quad (\text{A-96})$$

$$n_0 A_S = 2(\lambda_1 + \lambda_2 \sin \tau + \lambda_3 \cos \tau) \quad (\text{A-97})$$

$$n_0 A_R = -\lambda_2 \cos \tau + \lambda_3 \sin \tau \quad (\text{A-98})$$

$$n_0 A_W = -\lambda_5 \sin \tau + \lambda_6 \cos \tau \quad (\text{A-99})$$

3. Integrated Euler-Lagrange Equations

$$\lambda_1 = \lambda_{10} - \frac{3}{2} \lambda_4 \tau \quad (\text{A-100})$$

$$\lambda_2 = \lambda_{20} - 2\lambda_4 \cos \tau \quad (\text{A-101})$$

$$\lambda_3 = \lambda_{30} + 2\lambda_4 \sin \tau \quad (\text{A-102})$$

$$\lambda_4 = \text{CONSTANT} \quad (\text{A-103})$$

$$\lambda_5 = " \quad (\text{A-104})$$

$$\lambda_6 = " \quad (\text{A-105})$$

4. Boundary Conditions

A great simplification in the complexity of the equations can be achieved by taking advantage of the symmetry afforded by the Lagrange variables  $x_s$  and

$x_3$ . Therefore, in performing the integrations it will be convenient to use limits  $-\tau_f/2$  to  $\tau_f/2$  for the "in-plane" state variables.

State Variable ("in-plane")	<u>Transfer</u>		<u>Rendezvous</u>	
	$\tau = -\frac{\tau_f}{2}$	$\tau = \frac{\tau_f}{2}$	$\tau = -\frac{\tau_f}{2}$	$\tau = \frac{\tau_f}{2}$
$x_1$	1	$\Delta x_{1f} + 1$	1	$\Delta x_{1f} + 1$
$x_2$	$x_{20}$	$x_{20} + \Delta x_{2f}$	$x_{20}$	$x_{20} + \Delta x_{2f}$
$x_3$	$x_{30}$	$x_{30} + \Delta x_{3f}$	$x_{30}$	$x_{30} + \Delta x_{3f}$
$x_4$	$x_{40}$	FREE	$x_{40}$	$x_{40} + \Delta x_{4f}$
<u>(out-of-plane)</u>	$\tau = 0$	$\tau = \tau_f$	$\tau = 0$	$\tau = \tau_f$
$x_5$	0	$x_{5f}$	0	$x_{5f}$
$x_6$	0	$x_{6f}$	0	$x_{6f}$

### 5. Integrated Equations of State (with initial conditions)

$$\Delta x_1 = 4\lambda_{10}\left(\tau + \frac{\tau_f}{2}\right) - 4\lambda_{20}(\cos\tau - \cos\frac{\tau_f}{2}) + 4\lambda_{30}(\sin\tau + \sin\frac{\tau_f}{2}) - 3\lambda_4\left(\tau^2 - \frac{\tau_f^2}{4}\right) \quad (\text{A-106})$$

$$\begin{aligned} \Delta x_2 = & -4\lambda_{10}(\cos\tau - \cos\frac{\tau_f}{2}) + \frac{\lambda_{20}}{2}\left[5\left(\tau + \frac{\tau_f}{2}\right) - 3\left(\sin\tau \cos\tau + \frac{\sin\tau_f}{2}\right)\right] \\ & + \frac{3}{2}\lambda_{30}(\sin^2\tau - \sin^2\frac{\tau_f}{2}) - 2\lambda_4\left[4\left(\sin\tau + \sin\frac{\tau_f}{2}\right) - 3\left(\tau \cos\tau + \frac{\tau_f}{2} \cos\frac{\tau_f}{2}\right)\right] \end{aligned} \quad (\text{A-107})$$

$$\begin{aligned} \Delta x_3 = & 4\lambda_{10}(\sin\tau + \sin\frac{\tau_f}{2}) + \frac{3}{2}\lambda_{20}(\sin^2\tau - \sin^2\frac{\tau_f}{2}) \\ & + \frac{\lambda_{30}}{2}\left[5\left(\tau + \frac{\tau_f}{2}\right) + 3\left(\sin\tau \cos\tau + \frac{\sin\tau_f}{2}\right)\right] \\ & - 2\lambda_4\left[4\left(\cos\tau - \frac{\cos\tau_f}{2}\right) + 3\left(\tau \cos\tau + \frac{\tau_f}{2} \cos\frac{\tau_f}{2}\right)\right] \end{aligned} \quad (\text{A-108})$$

$$\begin{aligned}
\Delta x_4 = & \lambda_{10} \left\{ 3 \left( \tau + \frac{\tau_f}{2} \right)^2 - 8 \left[ 1 - \cos \left( \tau + \frac{\tau_f}{2} \right) \right] \right\} \\
& + \lambda_{20} \left\{ \left( \tau + \frac{\tau_f}{2} \right) \left[ 5 \cos \tau + 6 \cos \frac{\tau_f}{2} \right] - \frac{3}{2} \sin \left( \tau + \frac{\tau_f}{2} \right) - \frac{19}{2} \sin \tau - 8 \sin \frac{\tau_f}{2} \right\} \\
& + \lambda_{30} \left\{ \left( \tau + \frac{\tau_f}{2} \right) \left[ 6 \sin \frac{\tau_f}{2} - 5 \sin \tau \right] + \frac{3}{2} \cos \left( \tau + \frac{\tau_f}{2} \right) - \frac{19}{2} \cos \tau + 8 \cos \frac{\tau_f}{2} \right\} \\
& + \lambda_4 \left\{ 16 \left( \tau + \frac{\tau_f}{2} \right) - 6 \tau_f \left[ 1 - \cos \left( \tau + \frac{\tau_f}{2} \right) \right] + 3 \left( \frac{\tau_f}{2} \right)^3 + \frac{9}{2} \tau \left( \frac{\tau_f}{2} \right)^2 - \frac{3}{2} \tau^3 \right\} \\
& + 2 x_{20} \left[ \cos \tau - \cos \frac{\tau_f}{2} \right] - 2 x_{30} \left[ \sin \tau + \sin \frac{\tau_f}{2} \right]
\end{aligned} \tag{A-109}$$

$$x_5 = \frac{\lambda_5}{2} (\tau - \sin \tau \cos \tau) - \frac{\lambda_6}{2} \sin^2 \tau \tag{A-110}$$

$$x_6 = -\frac{\lambda_5}{2} \sin^2 \tau + \frac{\lambda_6}{2} (\tau + \sin \tau \cos \tau) \tag{A-111}$$

## 6. Transversality Conditions - Transfer

$$\lambda_4 = 0 \tag{A-112}$$

$$\frac{\lambda_5}{\lambda_6} = \tan \tau \tag{A-113}$$

## 7. Constants of Integration

### Transfer

$$\lambda_{10} = \frac{\frac{\Delta x_{1f}}{4} (5 \tau_f + 3 \sin \tau_f) - 4 \Delta x_{3f} \sin \frac{\tau_f}{2}}{\tau_f (5 \tau_f + 3 \sin \tau_f) - 16 (1 - \cos \tau_f)} \tag{A-114}$$

$$\lambda_{20} = \frac{2 \Delta x_{2f}}{5 \tau_f - 3 \sin \tau_f} \tag{A-115}$$

$$\lambda_{30} = \frac{2 \left[ \tau_f \Delta x_{3f} - 2 \Delta x_{1f} \sin \frac{\tau_f}{2} \right]}{\tau_f (5 \tau_f + 3 \sin \tau_f) - 16 (1 - \cos \tau_f)} \tag{A-116}$$

$$\lambda_5 = \frac{x_{5f} (\tau_f + \sin \tau_f \cos \tau_f) + x_{6f} \sin^2 \tau_f}{2(\tau_f^2 - \sin^2 \tau_f)} \quad (\text{A-117})$$

Rendezvous

$$\lambda_{10} = \frac{\frac{\Delta x_{1f}}{4} (5\tau_f + 3\sin \tau_f) - 4\Delta x_{3f} \sin \frac{\tau_f}{2}}{\tau_f (5\tau_f + 3\sin \tau_f) - 16(1 - \cos \tau_f)} \quad (\text{A-118})$$

$$\begin{aligned} \lambda_{20} = & \frac{1}{\tau_f (5\tau_f - 3\sin \tau_f) \left( \frac{3}{16} \tau_f^2 + 1 \right) - 2 \left( 8 \sin \frac{\tau_f}{2} - 3\tau_f \cos \frac{\tau_f}{2} \right)^2} \left[ \frac{3}{4} \tau_f \Delta x_{1f} \left( 3\tau_f \cos \frac{\tau_f}{2} - 8 \sin \frac{\tau_f}{2} \right) \right. \\ & + \Delta x_{2f} \left[ \frac{3}{8} \tau_f^3 + 8\tau_f - 3\tau_f(1 - \cos \tau_f) - 8\sin \tau_f \right] \\ & \left. - \left[ 3\tau_f \cos \frac{\tau_f}{2} - 8 \sin \frac{\tau_f}{2} \right] \left[ 2\Delta x_{3f} \sin \frac{\tau_f}{2} + \Delta x_{4f} + 4x_{30} \sin \frac{\tau_f}{2} \right] \right] \quad (\text{A-119}) \end{aligned}$$

$$\lambda_{30} = \frac{2 \left[ \tau_f \Delta x_{3f} - 2\Delta x_{1f} \sin \frac{\tau_f}{2} \right]}{\tau_f (5\tau_f + 3\sin \tau_f) - 16(1 - \cos \tau_f)} \quad (\text{A-120})$$

$$\begin{aligned} \lambda_4 = & \frac{1}{\tau_f (5\tau_f - 3\sin \tau_f) \left( \frac{3}{16} \tau_f^2 + 1 \right) - 2 \left( 8 \sin \frac{\tau_f}{2} - 3\tau_f \cos \frac{\tau_f}{2} \right)^2} \left[ -\frac{3}{16} \tau_f \Delta x_{1f} (5\tau_f - 3\sin \tau_f) \right. \\ & - \frac{\Delta x_{2f}}{2} \left[ 11\tau_f \cos \frac{\tau_f}{2} + 3 \sin \frac{\tau_f}{2} (1 - \cos \tau_f) - 22 \sin \frac{\tau_f}{2} \right] \\ & \left. + (5\tau_f - 3\sin \tau_f) \left[ \frac{\Delta x_{3f}}{2} \sin \frac{\tau_f}{2} + \frac{\Delta x_{4f}}{4} + x_{30} \sin \frac{\tau_f}{2} \right] \right] \quad (\text{A-121}) \end{aligned}$$

$$\lambda_5 = \frac{2 \left\{ x_{5f} (\tau_f + \sin \tau_f \cos \tau_f) + x_{6f} \sin^2 \tau_f \right\}}{\tau_f^2 - \sin^2 \tau_f} = \frac{2i \left[ \tau_f \sin \Omega_f + \sin \tau_f \sin(\Omega_f + \tau_f) \right]}{\tau_f^2 - \sin^2 \tau_f} \quad (\text{A-122})$$

$$\lambda_6 = \frac{2 \left\{ x_{5f} \sin^2 \tau_f + x_{6f} (\tau_f - \sin \tau_f \cos \tau_f) \right\}}{\tau_f^2 - \sin^2 \tau_f} = \frac{2i \left[ \tau_f \cos \Omega_f - \sin \tau_f \cos(\Omega_f + \tau_f) \right]}{\tau_f^2 - \sin^2 \tau_f} \quad (\text{A-123})$$

8. Controls

$$n_0 A_S = 2\lambda_{10} - 3\lambda_4 \tau + 2\lambda_{20} \sin \tau + 2\lambda_{30} \cos \tau \quad (A-124)$$

$$n_0 A_R = 2\lambda_4 - \lambda_{20} \cos \tau + \lambda_{30} \sin \tau \quad (A-125)$$

$$n_0 A_W = -\lambda_5 \sin \tau + \lambda_6 \cos \tau \quad (A-126)$$

9. PayoffTransfer

$$\frac{J}{n_0^3 r_0^2} = \frac{\frac{\Delta x_{1f}^2}{8} (5\tau_f + 3 \sin \tau_f) - 4 \Delta x_{1f} \Delta x_{3f} \sin \frac{\tau_f}{2} + \tau_f \Delta x_{3f}^2}{\tau_f (5\tau_f + 3 \sin \tau_f) - 16(1 - \cos \tau_f)} + \frac{\Delta x_{2f}^2}{5\tau_f - 3 \sin \tau_f} \quad (A-127)$$

$$+ \frac{i^2}{\tau_f + |\sin \tau_f|}$$

NOTE: (1)

Rendezvous

$$\frac{J}{n_0^3 r_0^2} = \frac{\frac{\Delta x_{1f}^2}{8} (5\tau_f + 3 \sin \tau_f) - 4 \Delta x_{1f} \Delta x_{3f} \sin \frac{\tau_f}{2} + \tau_f \Delta x_{3f}^2}{\tau_f (5\tau_f + 3 \sin \tau_f) - 16(1 - \cos \tau_f)}$$

$$+ \frac{\frac{1}{8} (5\tau_f - 3 \sin \tau_f) \left\{ 2 \Delta x_{2f} \cos \frac{\tau_f}{2} - 2 \Delta x_{3f} \sin \frac{\tau_f}{2} - \Delta x_{4f} + \frac{3}{4} \tau_f \Delta x_{1f} - 4 x_{30} \sin \frac{\tau_f}{2} \right\}^2}{\tau_f (5\tau_f - 3 \sin \tau_f) \left( \frac{3}{16} \tau_f^2 + 1 \right) - 2 \left( 3\tau_f \cos \frac{\tau_f}{2} - 8 \sin \frac{\tau_f}{2} \right)^2} \quad (A-128)$$

$$+ \frac{\Delta x_{2f} (3\tau_f \cos \frac{\tau_f}{2} - 8 \sin \frac{\tau_f}{2}) \left\{ 2 \Delta x_{2f} \cos \frac{\tau_f}{2} - 2 \Delta x_{3f} \sin \frac{\tau_f}{2} - \Delta x_{4f} + \frac{3}{4} \tau_f \Delta x_{1f} - 4 x_{30} \sin \frac{\tau_f}{2} \right\}}{\tau_f (5\tau_f - 3 \sin \tau_f) \left( \frac{3}{16} \tau_f^2 + 1 \right) - 2 \left( 3\tau_f \cos \frac{\tau_f}{2} - 8 \sin \frac{\tau_f}{2} \right)^2}$$

$$+ \frac{\tau_f \Delta x_{2f}^2 \left( \frac{3}{16} \tau_f^2 + 1 \right)}{\tau_f (5\tau_f - 3 \sin \tau_f) \left( \frac{3}{16} \tau_f^2 + 1 \right) - 2 \left( 3\tau_f \cos \frac{\tau_f}{2} - 8 \sin \frac{\tau_f}{2} \right)^2}$$

$$+ i^2 \left[ \frac{\tau_f - \sin \tau_f \cos (2\Omega_f + \tau_f)}{(\tau_f^2 - \sin^2 \tau_f)} \right]$$

NOTE: (1) The second term of this equation is incorrect in Ref. 1.

10. The optimal values for changes in the state variables  $x_4$  and  $\Omega$  are predicted by the variational analysis in the case of orbit transfer where the values  $x_4$  and  $\Omega$  are left open at the final time.

$$\begin{aligned} \Delta x_4^* = & \frac{3}{4} \tau_f \Delta x_{1f} - 2 \Delta x_{3f} \sin \frac{\tau_f}{2} - 4 x_{30} \sin \frac{\tau_f}{2} \\ & + \frac{4 \Delta x_{2f}}{5 \tau_f - 3 \sin \tau_f} \left\{ \frac{1}{2} \cos \frac{\tau_f}{2} (5 \tau_f - 3 \sin \tau_f) + 3 \tau_f \cos \frac{\tau_f}{2} - 8 \sin \frac{\tau_f}{2} \right\} \end{aligned} \quad (A-129)$$

$$\Omega_f^* = n \pi - \frac{\tau_f}{2} \quad (A-130)$$

#### 11. Payoff Equations with an Intermediate Reference Orbit

##### Transfer

$$\begin{aligned} \frac{J}{n_I^3 r_I^2} = & \frac{\frac{\Delta x_{1f}^2}{8} (5 \tau_f + 3 \sin \tau_f) - 4 \Delta x_{1f} \Delta x_{3f} \sin \frac{\tau_f}{2} + \tau_f \Delta x_{3f}^2}{\tau_f (5 \tau_f + 3 \sin \tau_f) - 16(1 - \cos \tau_f)} + \frac{\Delta x_{2f}^2}{5 \tau_f - 3 \sin \tau_f} \\ & + \frac{i^2}{\tau_f + |\sin \tau_f|} \end{aligned} \quad (A-131)$$

##### Rendezvous

$$\begin{aligned} \frac{J}{n_I^3 r_I^2} = & \frac{\frac{\Delta x_{1f}^2}{8} (5 \tau_f + 3 \sin \tau_f) - 4 \Delta x_{1f} \Delta x_{3f} \sin \frac{\tau_f}{2} + \tau_f \Delta x_{3f}^2}{\tau_f (5 \tau_f + 3 \sin \tau_f) - 16(1 - \cos \tau_f)} \\ & + \frac{\frac{1}{8} (5 \tau_f - 3 \sin \tau_f) \left\{ 2 \Delta x_{2f} \cos \frac{\tau_f}{2} - 2 \Delta x_{3f} \sin \frac{\tau_f}{2} - \Delta x_{4f} + \frac{3}{4} \tau_f (x_{10} + x_{1f} - 2) - 4 x_{30} \sin \frac{\tau_f}{2} \right\}^2}{\tau_f (5 \tau_f - 3 \sin \tau_f) \left( \frac{3}{16} \tau_f^2 + 1 \right) - 2 \left( 3 \tau_f \cos \frac{\tau_f}{2} - 8 \sin \frac{\tau_f}{2} \right)^2} \\ & + \frac{\Delta x_{2f} (3 \tau_f \cos \frac{\tau_f}{2} - 8 \sin \frac{\tau_f}{2}) \left\{ 2 \Delta x_{2f} \cos \frac{\tau_f}{2} - 2 \Delta x_{3f} \sin \frac{\tau_f}{2} - \Delta x_{4f} + \frac{3}{4} \tau_f (x_{10} + x_{1f} - 2) - 4 x_{30} \sin \frac{\tau_f}{2} \right\}}{\tau_f (5 \tau_f - 3 \sin \tau_f) \left( \frac{3}{16} \tau_f^2 + 1 \right) - 2 \left( 3 \tau_f \cos \frac{\tau_f}{2} - 8 \sin \frac{\tau_f}{2} \right)^2} \\ & + \frac{\tau_f \Delta x_{2f}^2 \left( \frac{3}{16} \tau_f^2 + 1 \right)}{\tau_f (5 \tau_f - 3 \sin \tau_f) \left( \frac{3}{16} \tau_f^2 + 1 \right) - 2 \left( 3 \tau_f \cos \frac{\tau_f}{2} - 8 \sin \frac{\tau_f}{2} \right)^2} \\ & + i^2 \left[ \frac{\tau_f - \sin \tau_f \cos (2 \Omega_f + \tau_f)}{\tau_f^2 - \sin^2 \tau_f} \right] \end{aligned} \quad (A-132)$$

12. Optimal Transfer Coordinates

$$\Delta x_4^* = \frac{3}{4} \tau_f (x_{10} + x_{1f} - 2) - 2\Delta x_{3f} \sin \frac{\tau_f}{2} - 4x_{30} \sin \frac{\tau_f}{2} \quad (A-133)$$

$$+ \frac{4 \Delta x_{2f}}{5\tau_f - 3\sin\tau_f} \left\{ \frac{1}{2} \cos \frac{\tau_f}{2} (5\tau_f - 3\sin\tau_f) + 3\tau_f \cos \frac{\tau_f}{2} - 8 \sin \frac{\tau_f}{2} \right\}$$

$$\Omega_f^* = n\pi - \frac{\tau_f}{2} \quad (A-134)$$

## APPENDIX III

## SYNTHESIS OF THE OPTIMAL CONTROLS

A. Rotating Coordinates1. Control Equations

$$A_y = \frac{\partial A_y}{\partial y} y + \frac{\partial A_y}{\partial u} u + \frac{\partial A_y}{\partial v} v + \frac{\partial A_y}{\partial x} x \quad (A-135)$$

$$A_x = \frac{\partial A_x}{\partial y} y + \frac{\partial A_x}{\partial u} u + \frac{\partial A_x}{\partial v} v + \frac{\partial A_x}{\partial x} x \quad (A-136)$$

$$A_z = \frac{\partial A_z}{\partial z} z + \frac{\partial A_z}{\partial w} w \quad (A-137)$$

2. Guidance Coefficients - Transfer

$$\frac{\partial A_y}{\partial y} = \frac{12 \tau'}{\Phi} (1 - \cos \tau') (29 - 27 \cos \tau') \quad (A-138)$$

$$\frac{\partial A_y}{\partial u} = \frac{24}{\Phi} (1 - \cos \tau') (11 \sin \tau' - 3 \tau' \cos \tau' - 8 \tau') \quad (A-139)$$

$$\frac{\partial A_y}{\partial v} = \frac{12}{\Phi} (5 \tau'^2 + 3 \tau' \sin \tau' \cos \tau' - 8 \sin^2 \tau') \quad (A-140)$$

$$\frac{\partial A_x}{\partial y} = \frac{12}{\Phi} \left[ 70 \tau' \sin \tau' - 55 \tau'^2 + 18 \tau' \sin \tau' \cos \tau' + 3(1 - \cos \tau')(5 - 27 \cos \tau') \right] \quad (A-141)$$

$$\frac{\partial A_x}{\partial u} = \frac{6}{\Phi} \left[ 65 \tau'^2 - 80 \tau' \sin \tau' - 24 \tau' \sin \tau' \cos \tau' - (1 - \cos \tau')(25 - 103 \cos \tau') \right] \quad (A-142)$$

$$\frac{\partial A_x}{\partial v} = - \frac{24}{\Phi} (8 \tau' - 11 \sin \tau' + 3 \tau' \cos \tau')(1 - \cos \tau')^{(1)} \quad (A-143)$$

$$^{(1)} \text{NOTE : } \frac{\partial A_x}{\partial v} = \frac{\partial A_y}{\partial u}$$



$$\frac{\partial A_z}{\partial z} = \frac{-2 \sin^2 \tau'}{\tau'^2 - \sin^2 \tau'} \quad (\text{A-144})$$

$$\frac{\partial A_z}{\partial w} = \frac{-(2\tau' - \sin 2\tau')}{\tau'^2 - \sin^2 \tau'} \quad (\text{A-145})$$

where

$$\Phi = 480\tau' - 75\tau'^3 - 240\tau'\cos\tau'(1 + \cos\tau') - 144\sin\tau'(1 - \cos\tau') - 213\tau'\sin^2\tau' \quad (\text{A-146})$$

### 3. Rendezvous

Due to the length and complexity of the synthesized, in-plane, control equations for rendezvous, the guidance coefficients are not written explicitly here. Instead the basic equations are tabulated, and the coefficients calculated from these equations are plotted in Figs. 20 through 22.

$$\frac{\partial A_x}{\partial x_i} = 3 \frac{\partial C_4}{\partial x_i} - 3 \frac{\partial C_0}{\partial x_i} \tau' - 4 \frac{\partial C_1}{\partial x_i} \cos\tau' - 4 \frac{\partial C_2}{\partial x_i} \sin\tau' \quad (\text{A-147})$$

$$\frac{\partial A_y}{\partial x_i} = 2 \left( \frac{\partial C_0}{\partial x_i} - \frac{\partial C_1}{\partial x_i} \sin\tau' + \frac{\partial C_2}{\partial x_i} \cos\tau' \right) \quad (\text{A-148})$$

$$A_z = \frac{\partial A_z}{\partial z} z + \frac{\partial A_z}{\partial w} w \quad (\text{A-149})$$

$$C_0 = \begin{vmatrix} x & \phi_{11} & \phi_{12} & \phi_{14} \\ y & \phi_{21} & \phi_{22} & \phi_{24} \\ u & \phi_{31} & \phi_{32} & \phi_{34} \\ v & \phi_{41} & \phi_{42} & \phi_{44} \end{vmatrix} \quad \text{D} \quad (\text{A-150})$$

$$C_1 = \begin{vmatrix} \phi_{10} & x & \phi_{12} & \phi_{14} \\ \phi_{20} & y & \phi_{22} & \phi_{24} \\ \phi_{30} & u & \phi_{32} & \phi_{34} \\ \phi_{40} & v & \phi_{42} & \phi_{44} \end{vmatrix} \quad \text{D} \quad (\text{A-151})$$

$$C_2 = \begin{vmatrix} \phi_{10} & \phi_{11} & x & \phi_{14} \\ \phi_{20} & \phi_{21} & y & \phi_{24} \\ \phi_{30} & \phi_{31} & u & \phi_{34} \\ \phi_{40} & \phi_{41} & v & \phi_{44} \end{vmatrix}$$

D

(A-152)

$$C_4 = \begin{vmatrix} \phi_{10} & \phi_{11} & \phi_{12} & x \\ \phi_{20} & \phi_{21} & \phi_{22} & y \\ \phi_{30} & \phi_{31} & \phi_{32} & u \\ \phi_{40} & \phi_{41} & \phi_{42} & v \end{vmatrix}$$

D

(A-153)

where

$$D = \begin{vmatrix} \phi_{10} & \phi_{11} & \phi_{12} & \phi_{14} \\ \phi_{20} & \phi_{21} & \phi_{22} & \phi_{24} \\ \phi_{30} & \phi_{31} & \phi_{32} & \phi_{34} \\ \phi_{40} & \phi_{41} & \phi_{42} & \phi_{44} \end{vmatrix}$$

(A-154)

and

$$\begin{aligned} \phi_{10} &= \frac{3}{4} \tau'^3 - 8\tau' + 8\sin\tau' & \phi_{30} &= 8(1 - \cos\tau') - \frac{9}{4} \tau'^2 \\ \phi_{11} &= 8(1 - \cos\tau') - 5\tau'\sin\tau' & \phi_{31} &= 5\tau'\cos\tau' - 3\sin\tau' \\ \phi_{12} &= 5\tau'\cos\tau' - 11\sin\tau' + 6\tau' & \phi_{32} &= 5\tau'\sin\tau' - 6(1 - \cos\tau') \\ \phi_{14} &= 6(1 - \cos\tau') - \frac{9}{4} \tau'^2 & \phi_{34} &= \frac{9}{2} \tau' - 6\sin\tau' \\ \phi_{20} &= 4(1 - \cos\tau') - \frac{3}{2} \tau'^2 & \phi_{40} &= 3\tau' - 4\sin\tau' \\ \phi_{21} &= \frac{5}{2} [\tau'\cos\tau' - \sin\tau'] & \phi_{41} &= \frac{5}{2} \tau'\sin\tau' \\ \phi_{22} &= \frac{5}{2} \tau'\sin\tau' - 4(1 - \cos\tau') & \phi_{42} &= \frac{3}{2} \sin\tau' - \frac{5}{2} \tau'\cos\tau' \\ \phi_{24} &= 3(\tau' - \sin\tau') & \phi_{44} &= -3(1 - \cos\tau') \end{aligned}$$

(A-155)

$$\frac{\partial A_z}{\partial z} = \frac{-2 \sin^2 \tau'}{\tau'^2 - \sin^2 \tau'} \quad (\text{A-156})$$

$$\frac{\partial A_z}{\partial w} = \frac{-(2\tau' - \sin 2\tau')}{\tau'^2 - \sin^2 \tau'} \quad (\text{A-157})$$

## B. Lagrange Variables

### 1. Control Equations

$$A_R = \frac{\partial A_R}{\partial \Delta x_1} \Delta x_1 + \frac{\partial A_R}{\partial \Delta x_2} \Delta x_2 + \frac{\partial A_R}{\partial \Delta x_3} \Delta x_3 + \frac{\partial A_R}{\partial \Delta x_4} \Delta x_4 + \frac{\partial A_R}{\partial x_{30}} x_{30} \quad (\text{A-158})$$

$$A_S = \frac{\partial A_S}{\partial \Delta x_1} \Delta x_1 + \frac{\partial A_S}{\partial \Delta x_2} \Delta x_2 + \frac{\partial A_S}{\partial \Delta x_3} \Delta x_3 + \frac{\partial A_S}{\partial \Delta x_4} \Delta x_4 + \frac{\partial A_S}{\partial x_{30}} x_{30} \quad (\text{A-159})$$

$$A_W = \frac{\partial A_W}{\partial \Delta x_5} \Delta x_5 + \frac{\partial A_W}{\partial \Delta x_6} \Delta x_6 \quad (\text{A-160})$$

### 2. Guidance Coefficients - Transfer

$$\frac{\partial A_R}{\partial \Delta x_1} = \frac{-4 \sin \tau' \sin \frac{\tau'}{2}}{\tau'(5\tau' + 3 \sin \tau') - 16(1 - \cos \tau')} \quad (\text{A-161})$$

$$\frac{\partial A_R}{\partial \Delta x_2} = \frac{2 \cos \tau'}{5\tau' - 3 \sin \tau'} \quad (\text{A-162})$$

$$\frac{\partial A_R}{\partial \Delta x_3} = \frac{2\tau' \sin \tau'}{\tau'(5\tau' + 3 \sin \tau') - 16(1 - \cos \tau')} \quad (\text{A-163})$$

$$\frac{\partial A_S}{\partial \Delta x_1} = \frac{8 \cos \tau' \sin \frac{\tau'}{2} - \frac{1}{2}(5\tau' + 3 \sin \tau')}{\tau'(5\tau' + 3 \sin \tau') - 16(1 - \cos \tau')} \quad (\text{A-164})$$

$$\frac{\partial A_S}{\partial \Delta x_2} = \frac{4 \sin \tau'}{5\tau' - 3 \sin \tau'} \quad (\text{A-165})$$

$$\frac{\partial A_W}{\partial \Delta x_9} = \frac{\cos \tau' \sin^2 \tau'}{\tau'^2 - \sin^2 \tau'} \quad (A-166)$$

$$\frac{\partial A_W}{\partial \Delta x_8} = \frac{-\frac{1}{2} \cos \tau' (2\tau' - \sin 2\tau')}{\tau'^2 - \sin^2 \tau'} \quad (A-167)$$

$$\frac{\partial A_S}{\partial \Delta x_3} = \frac{4(2 \sin \frac{\tau'}{2} - \tau' \cos \tau')}{\tau'(5\tau' + 3 \sin \tau') - 16(1 - \cos \tau')} \quad (A-168)$$

### 3. Guidance Coefficients - Rendezvous

$$\frac{\partial A_R}{\partial \Delta x_1} = \frac{4 \sin \tau' \sin \frac{\tau'}{2}}{Q} - \frac{\frac{3}{8} \tau' [5\tau' - 3 \sin \tau' + 2 \cos \tau' (3\tau' \cos \frac{\tau'}{2} - 8 \sin \frac{\tau'}{2})]}{B} \quad (A-169)$$

$$\frac{\partial A_R}{\partial \Delta x_2} = \frac{2\tau' \cos \tau' (\frac{3}{16} \tau'^2 + 1) + \cos \frac{\tau'}{2} (5\tau' - 3 \sin \tau') + 2(3\tau' \cos \frac{\tau'}{2} - 8 \sin \frac{\tau'}{2})(1 + \cos \tau' \cos \frac{\tau'}{2})}{B} \quad (A-170)$$

$$\frac{\partial A_R}{\partial \Delta x_3} = \frac{-2\tau' \sin \tau' + \sin \frac{\tau'}{2} [5\tau' - 3 \sin \tau' + 2 \cos \tau' (3\tau' \cos \frac{\tau'}{2} - 8 \sin \frac{\tau'}{2})]}{B} \quad (A-171)$$

$$\frac{\partial A_R}{\partial \Delta x_4} = -\frac{\frac{1}{2} [5\tau' - 3 \sin \tau' + 2 \cos \tau' (3\tau' \cos \frac{\tau'}{2} - 8 \sin \frac{\tau'}{2})]}{B} \quad (A-172)$$

$$\frac{\partial A_R}{\partial \Delta x_{30}} = \frac{2 \sin \frac{\tau'}{2} [5\tau' - 3 \sin \tau' + 2 \cos \tau' (3\tau' \cos \frac{\tau'}{2} - 8 \sin \frac{\tau'}{2})]}{B} \quad (A-173)$$

$$\frac{\partial A_S}{\partial \Delta x_1} = \frac{\frac{1}{2} (5\tau' + 3 \sin \tau' - 16 \sin \frac{\tau'}{2} \cos \tau')}{Q} - \frac{\frac{3}{16} \tau' [3\tau' (5\tau' - 3 \sin \tau') + 8 \sin \tau' (3\tau' \cos \frac{\tau'}{2} - 8 \sin \frac{\tau'}{2})]}{B} \quad (A-174)$$

$$\frac{\partial A_S}{\partial \Delta x_2} = \frac{4\tau' \sin \tau' \left( \frac{3}{16} \tau'^2 + 1 \right) + \frac{3}{2} \tau' \cos \frac{\tau'}{2} (5\tau' - 3 \sin \tau')}{(3\tau' \cos \frac{\tau'}{2} - 8 \sin \frac{\tau'}{2})(3\tau' + 4 \sin \tau' \cos \frac{\tau'}{2})} + \frac{B}{B} \quad (A-175)$$

$$\frac{\partial A_S}{\partial \Delta x_3} = \frac{4(\tau' \cos \tau' - 2 \sin \frac{\tau'}{2})}{\frac{1}{2} \sin \frac{\tau'}{2} \left[ 3\tau'(5\tau' - 3 \sin \tau') + 8 \sin \tau' (3\tau' \cos \frac{\tau'}{2} - 8 \sin \frac{\tau'}{2}) \right]} + \frac{B}{B} \quad (A-176)$$

$$\frac{\partial A_S}{\partial \Delta x_4} = - \frac{\frac{1}{4} \left[ 3\tau'(5\tau' - 3 \sin \tau') + 8 \sin \tau' (3\tau' \cos \frac{\tau'}{2} - 8 \sin \frac{\tau'}{2}) \right]}{B} \quad (A-177)$$

$$\frac{\partial A_S}{\partial x_{30}} = \frac{\sin \frac{\tau'}{2} \left[ 3\tau'(5\tau' - 3 \sin \tau') + 8 \sin \tau' (3\tau' \cos \frac{\tau'}{2} - 8 \sin \frac{\tau'}{2}) \right]}{B} \quad (A-178)$$

$$\frac{\partial A_W}{\partial \Delta x_5} = \frac{2 \sin \tau' (\tau' + \sin 2\tau')}{\tau'^2 - \sin^2 \tau'} \quad (A-179)$$

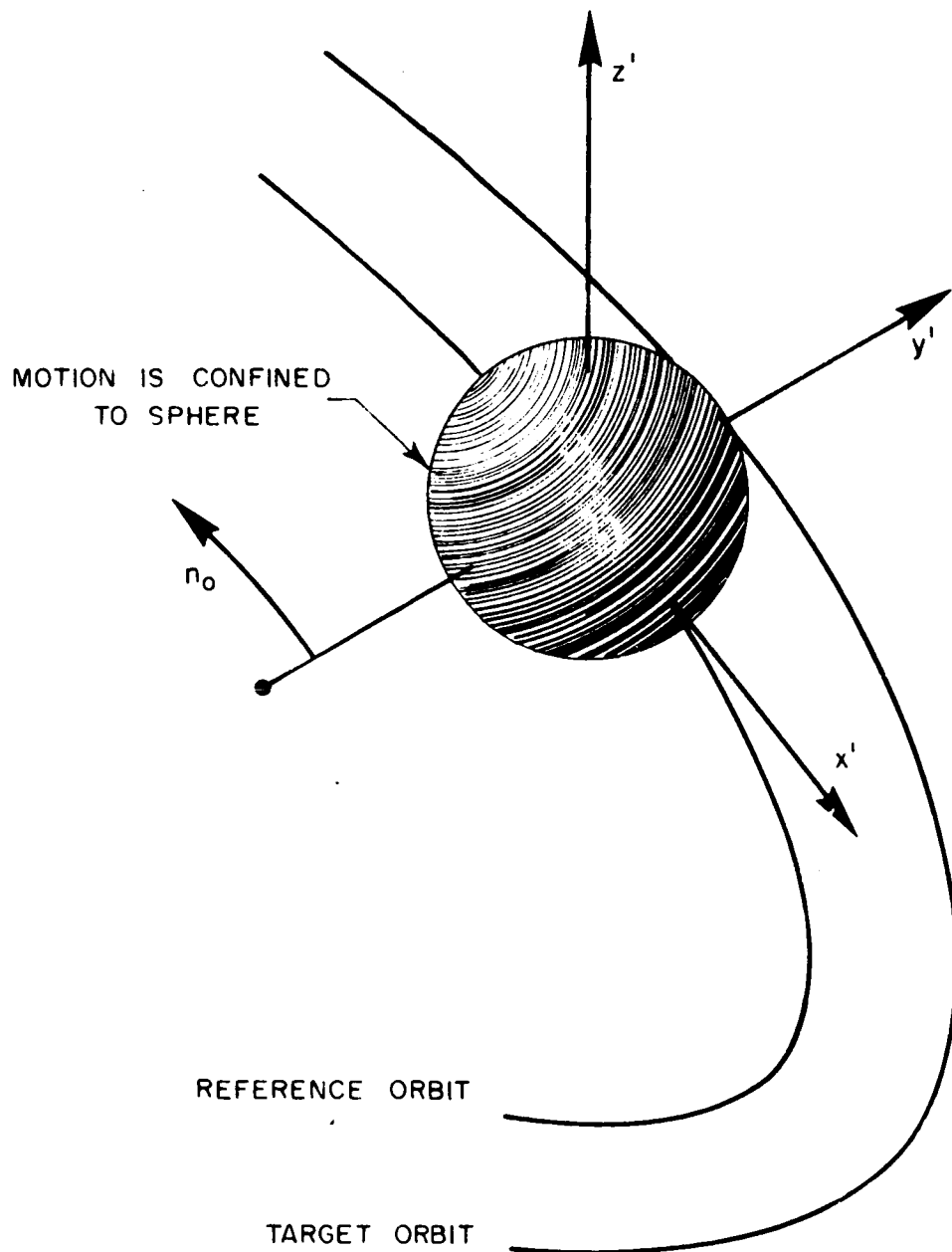
$$\frac{\partial A_W}{\partial \Delta x_6} = - \frac{2 \left[ \sin \tau' + \cos \tau' (\tau' - \sin 2\tau') \right]}{\tau'^2 - \sin^2 \tau'} \quad (A-180)$$

where

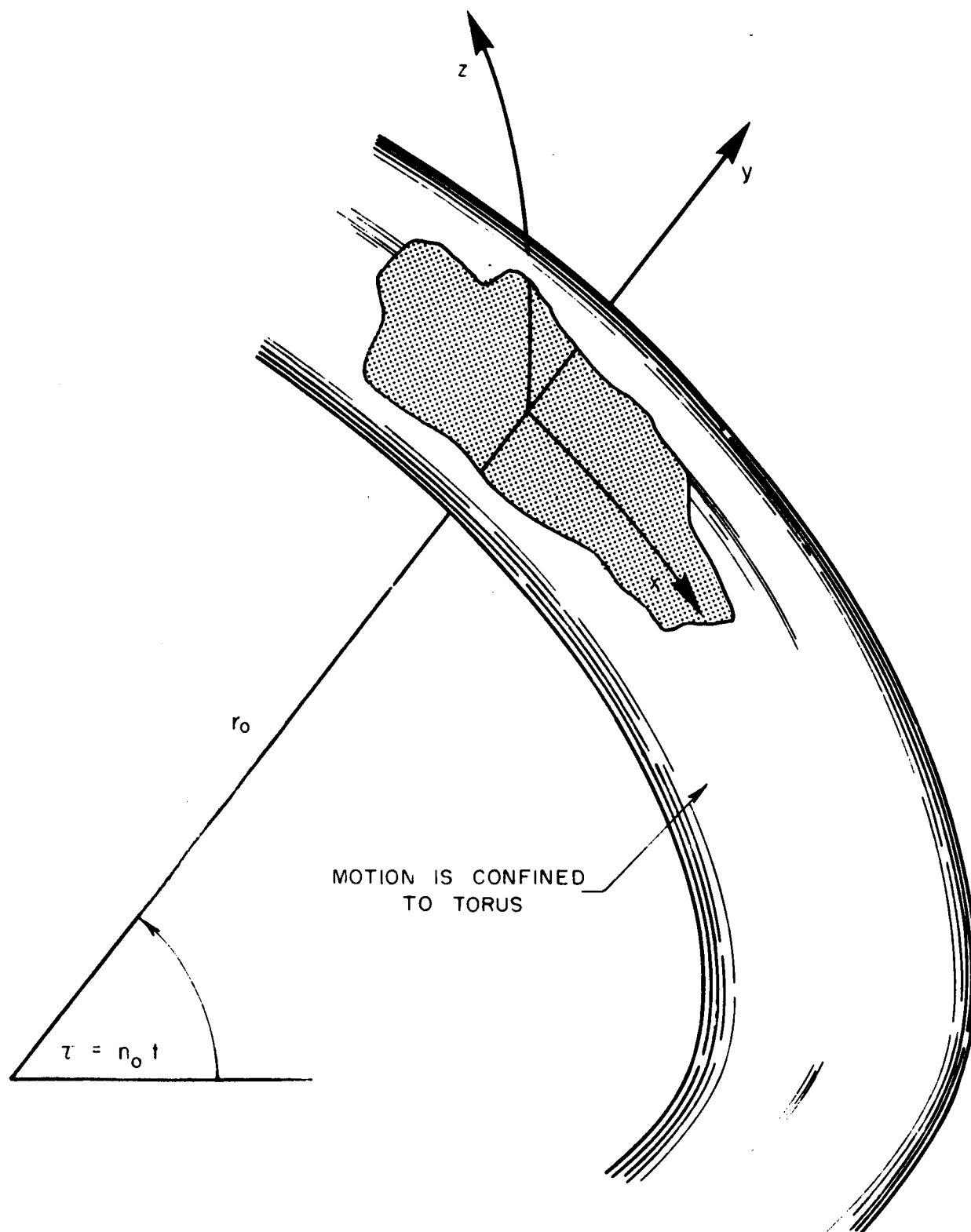
$$Q = 16(1 - \cos \tau') - \tau'(5\tau' + 3 \sin \tau') \quad (A-181)$$

$$B = \tau'(5\tau' - 3 \sin \tau') \left( \frac{3}{16} \tau'^2 + 1 \right) - 2 \left( 8 \sin \frac{\tau'}{2} - 3\tau' \cos \frac{\tau'}{2} \right)^2 \quad (A-182)$$

## RECTANGULAR COORDINATE SYSTEM

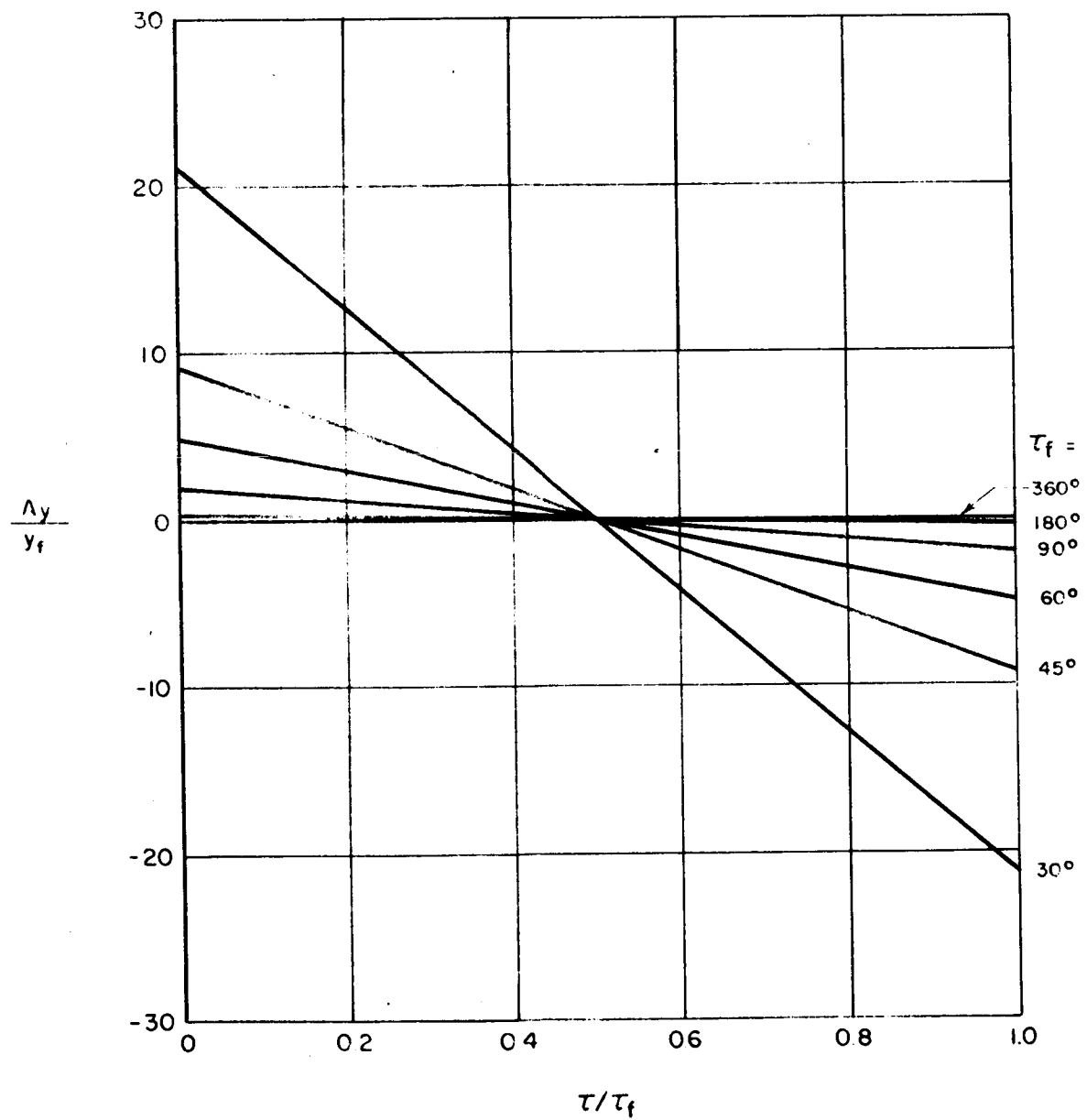


## SPHERICAL COORDINATE SYSTEM



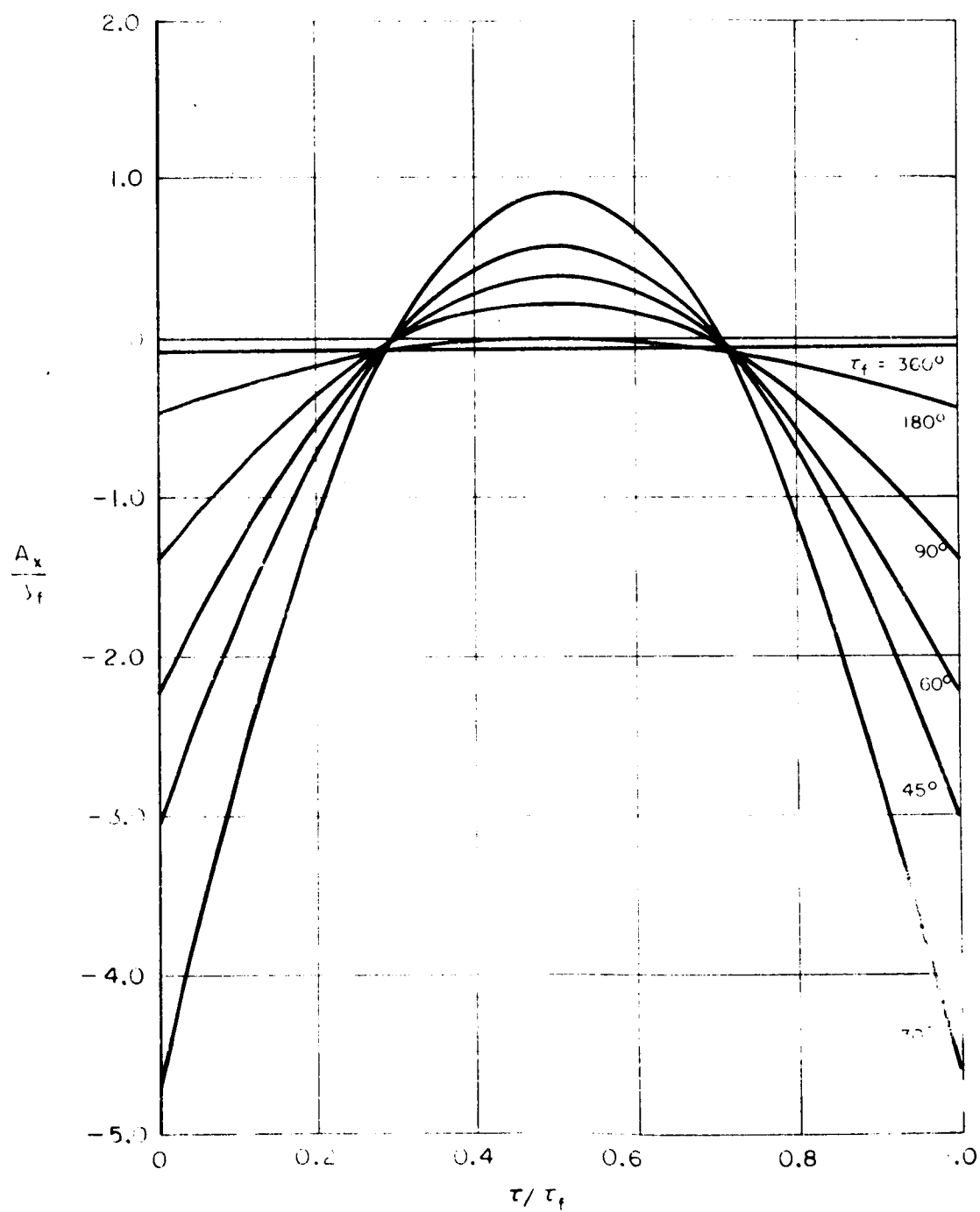
## RADIAL ACCELERATION

## CIRCLE-TO-CIRCLE TRANSFER

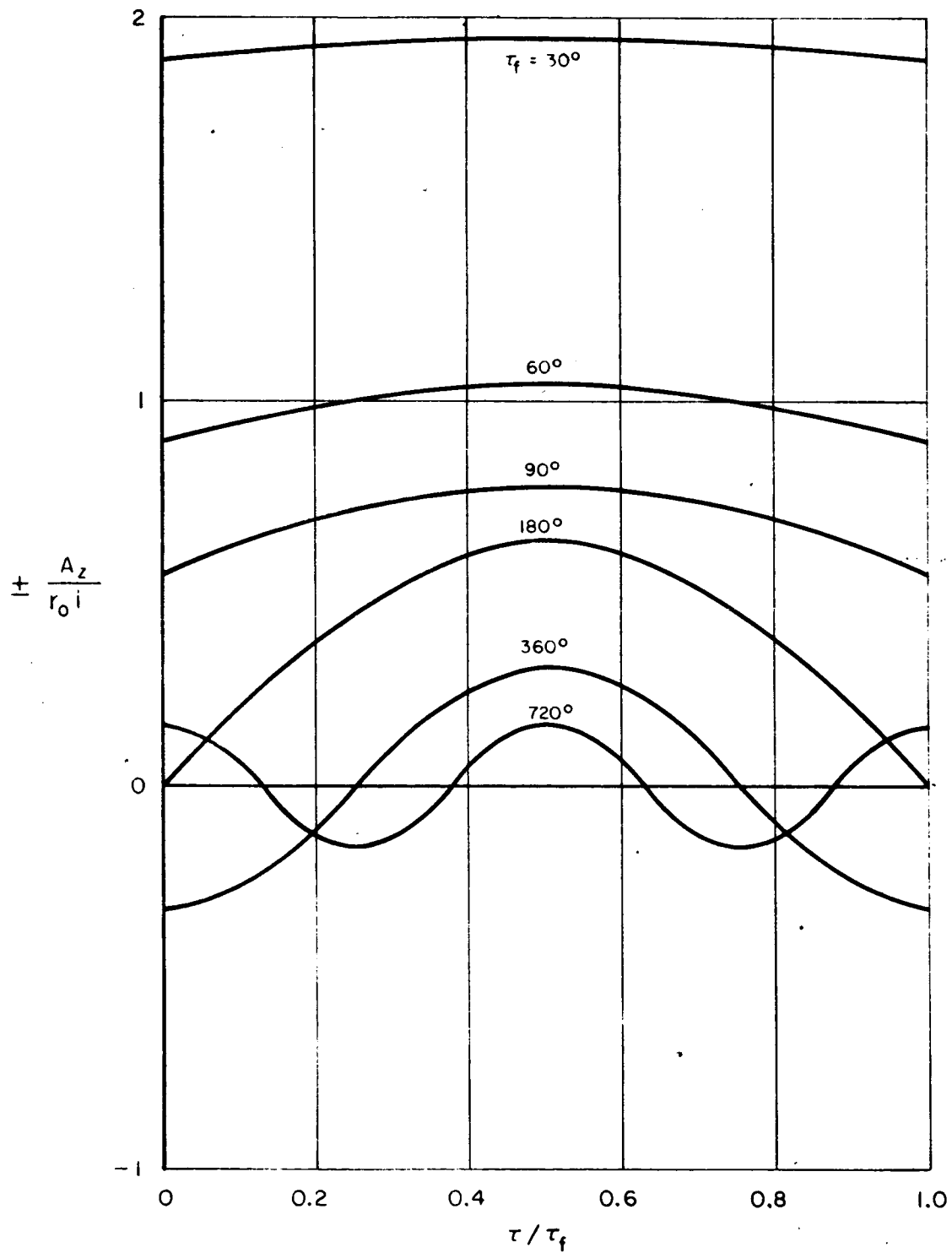




CIRCUMFERENTIAL ACCELERATION  
CIRCLE - TO - CIRCLE TRANSFER

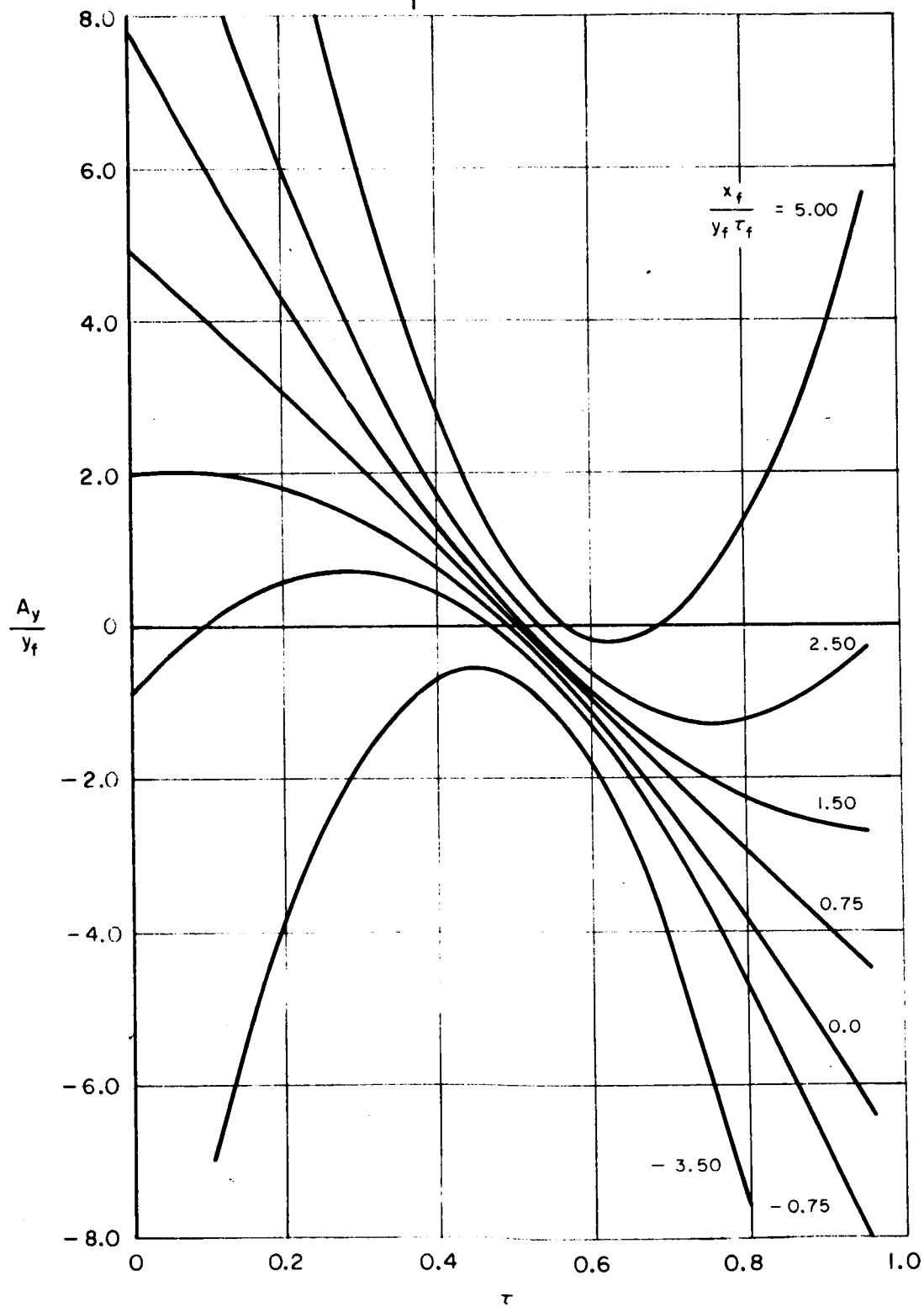


NORMAL ACCELERATION  
CIRCLE-TO-CIRCLE TRANSFER



RADIAL ACCELERATION  
CIRCLE - TO - CIRCLE RENDEZVOUS

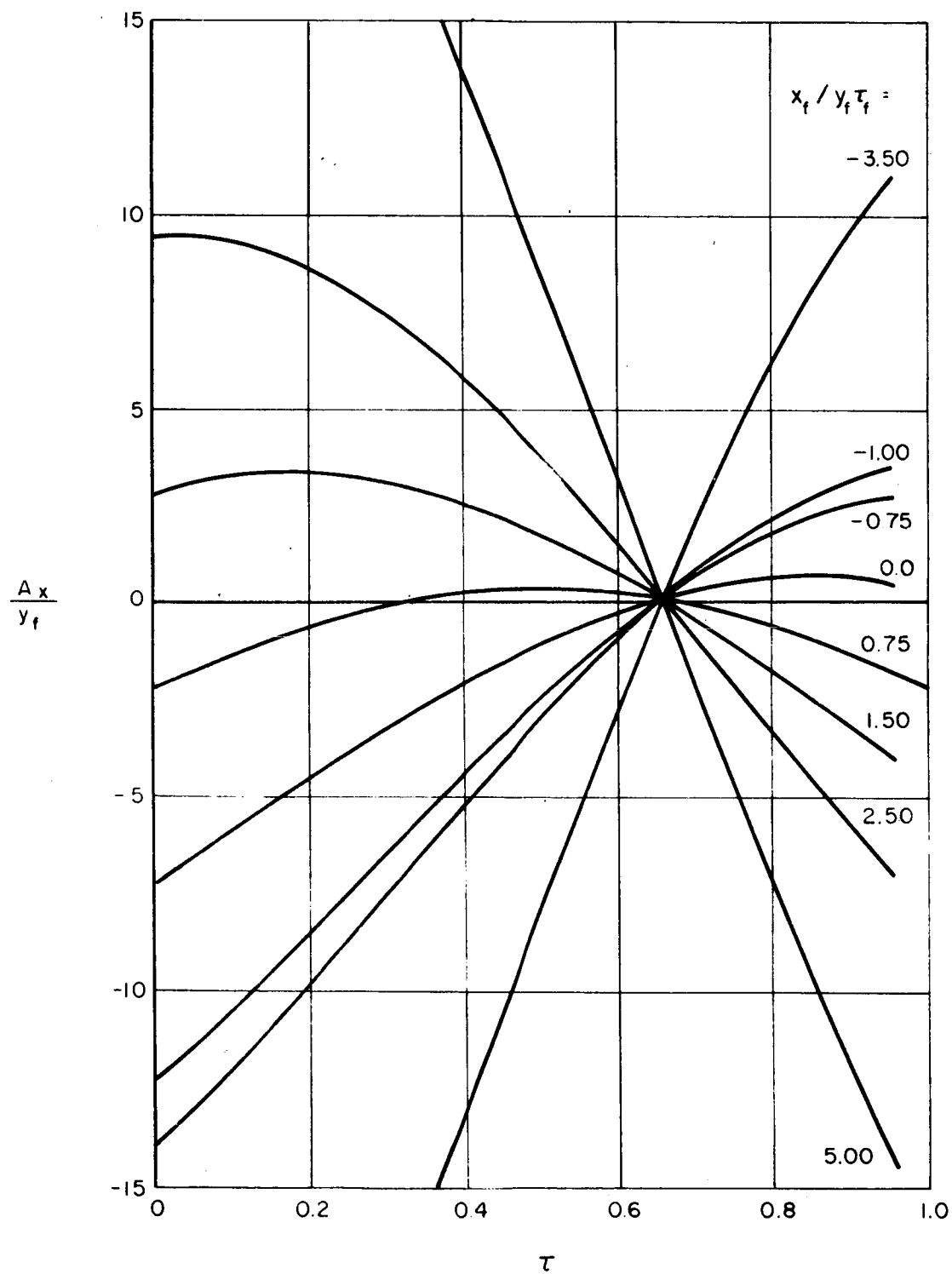
$$\tau_f = 60^\circ$$



## CIRCUMFERENTIAL ACCELERATION

CIRCLE - TO - CIRCLE RENDEZVOUS

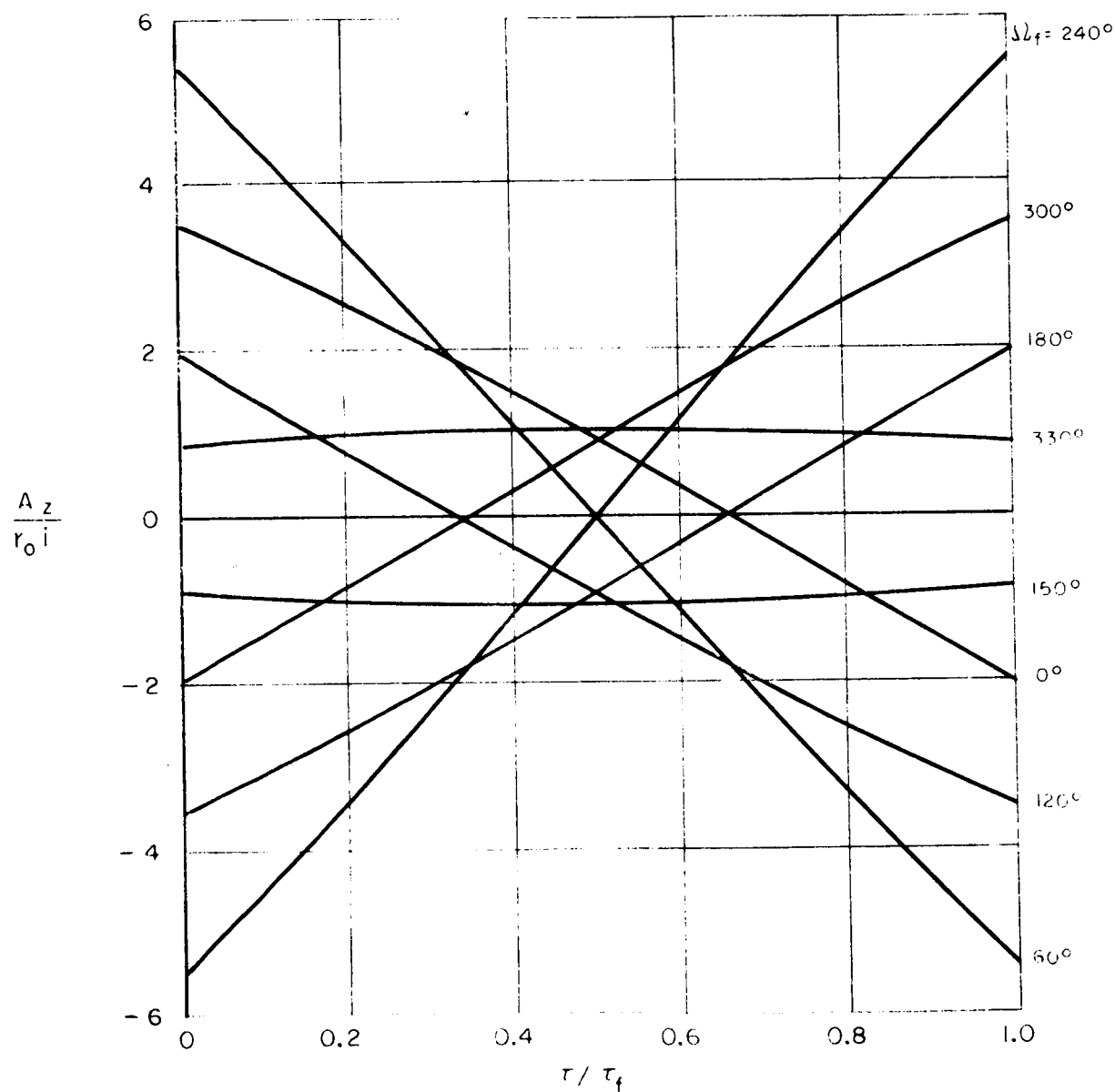
$$\tau_f = 60^\circ$$



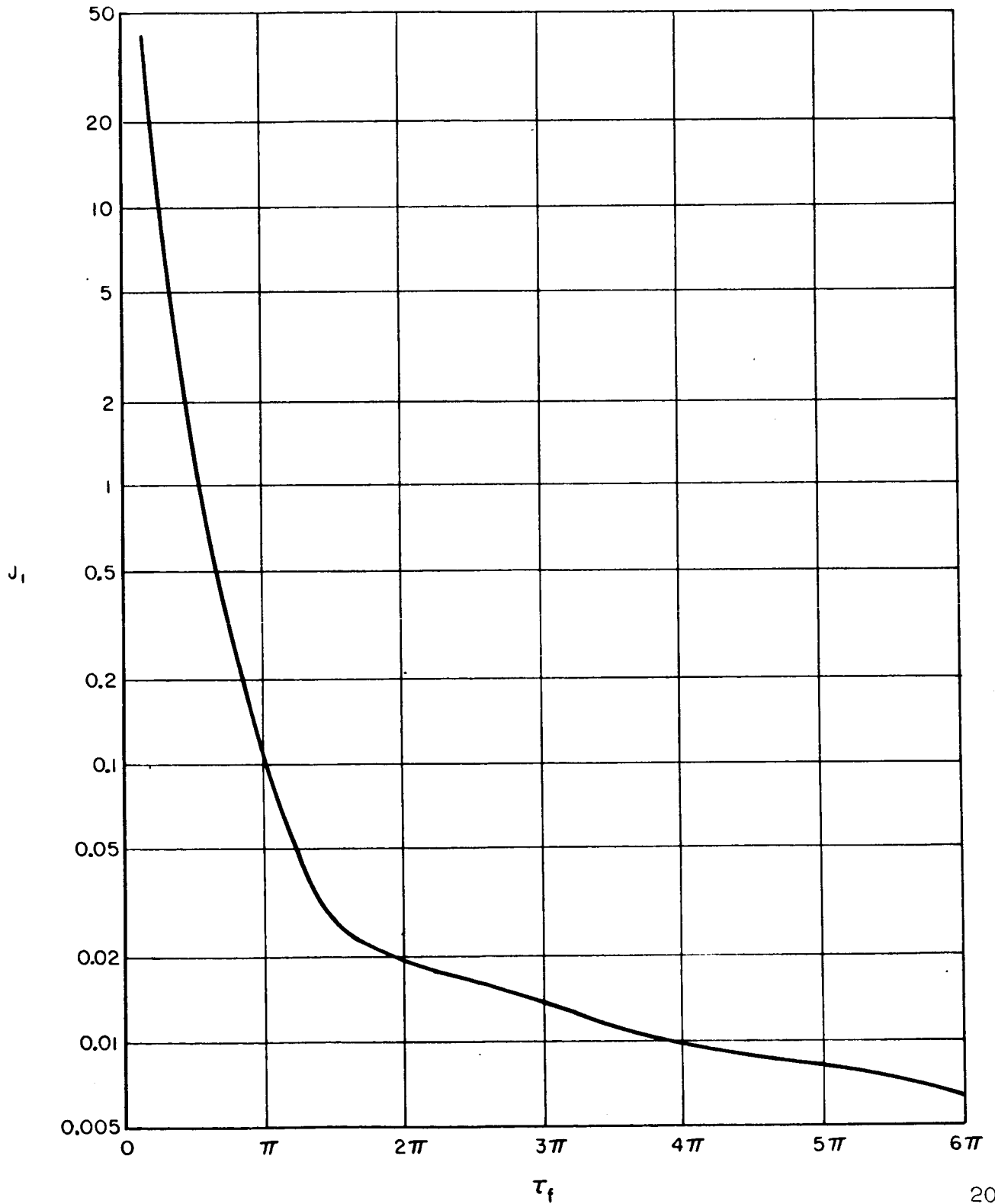
## NORMAL ACCELERATION

CIRCLE - TO - CIRCLE RENDEZVOUS

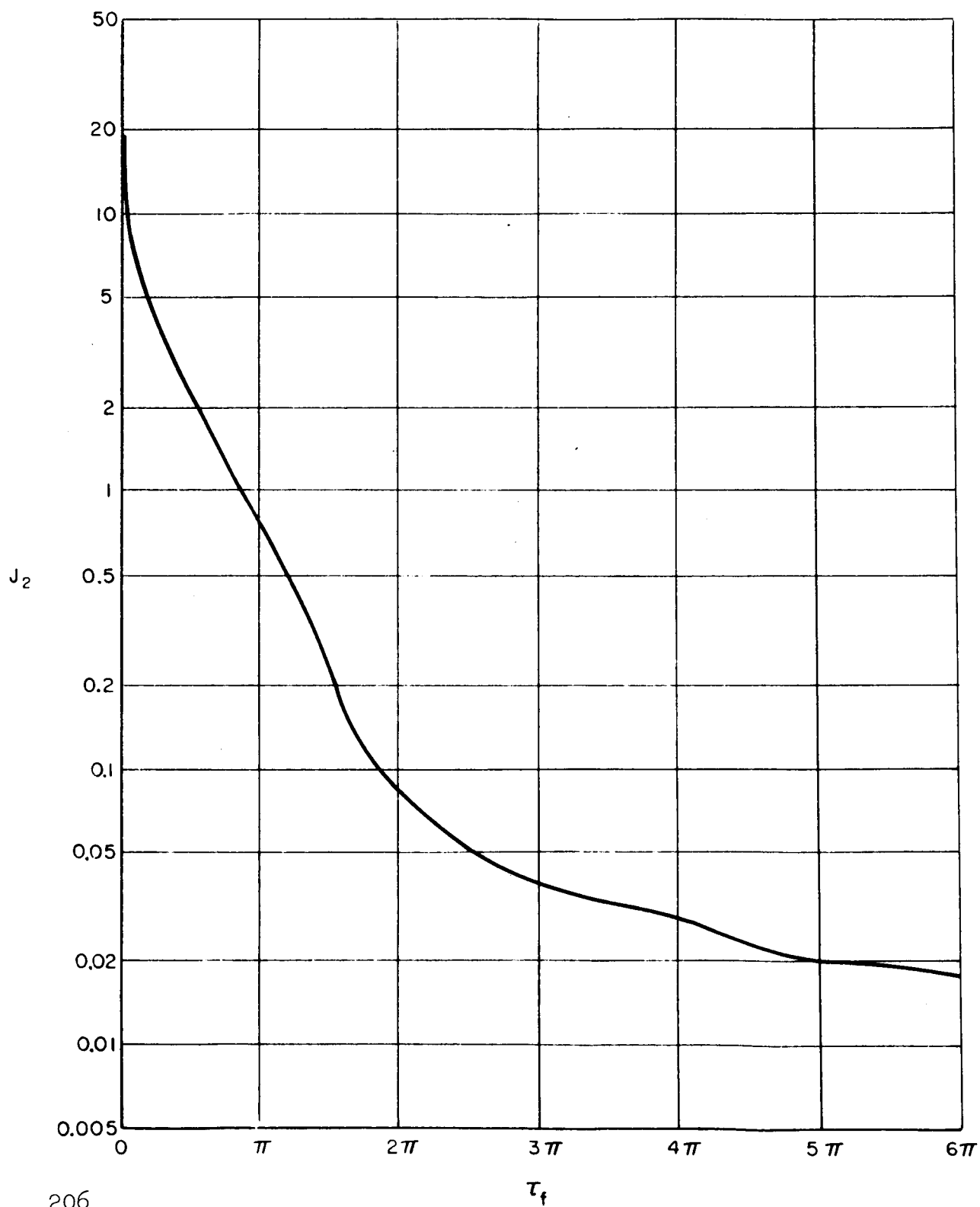
$$\tau_f = 60^\circ$$

OPTIMUM TRANSFERS :  $\Omega_f = 150^\circ, 330^\circ$ 

IN-PLANE COMPONENT OF J  
CIRCLE-TO-CIRCLE TRANSFER



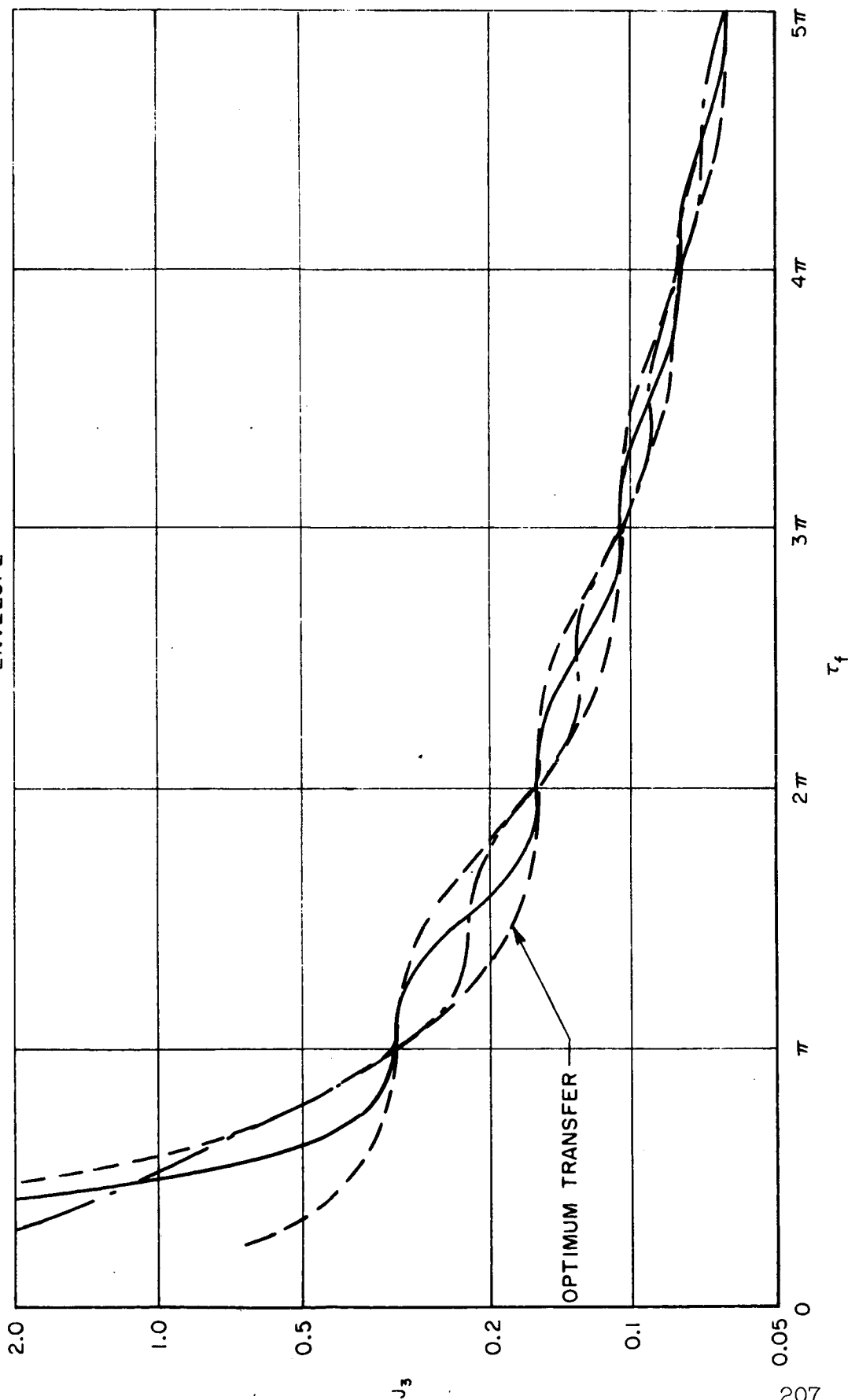
IN-PLANE COMPONENT OF J  
CIRCLE-TO-CIRCLE RENDEZVOUS



## OUT-OF-PLANE COMPONENT OF J

CIRCLE - TO - CIRCLE RENDEZVOUS

$\Omega_f = 0, \pi$   
 $\Omega_f = \frac{\pi}{2}, \frac{3\pi}{2}$   
 ENVELOPE

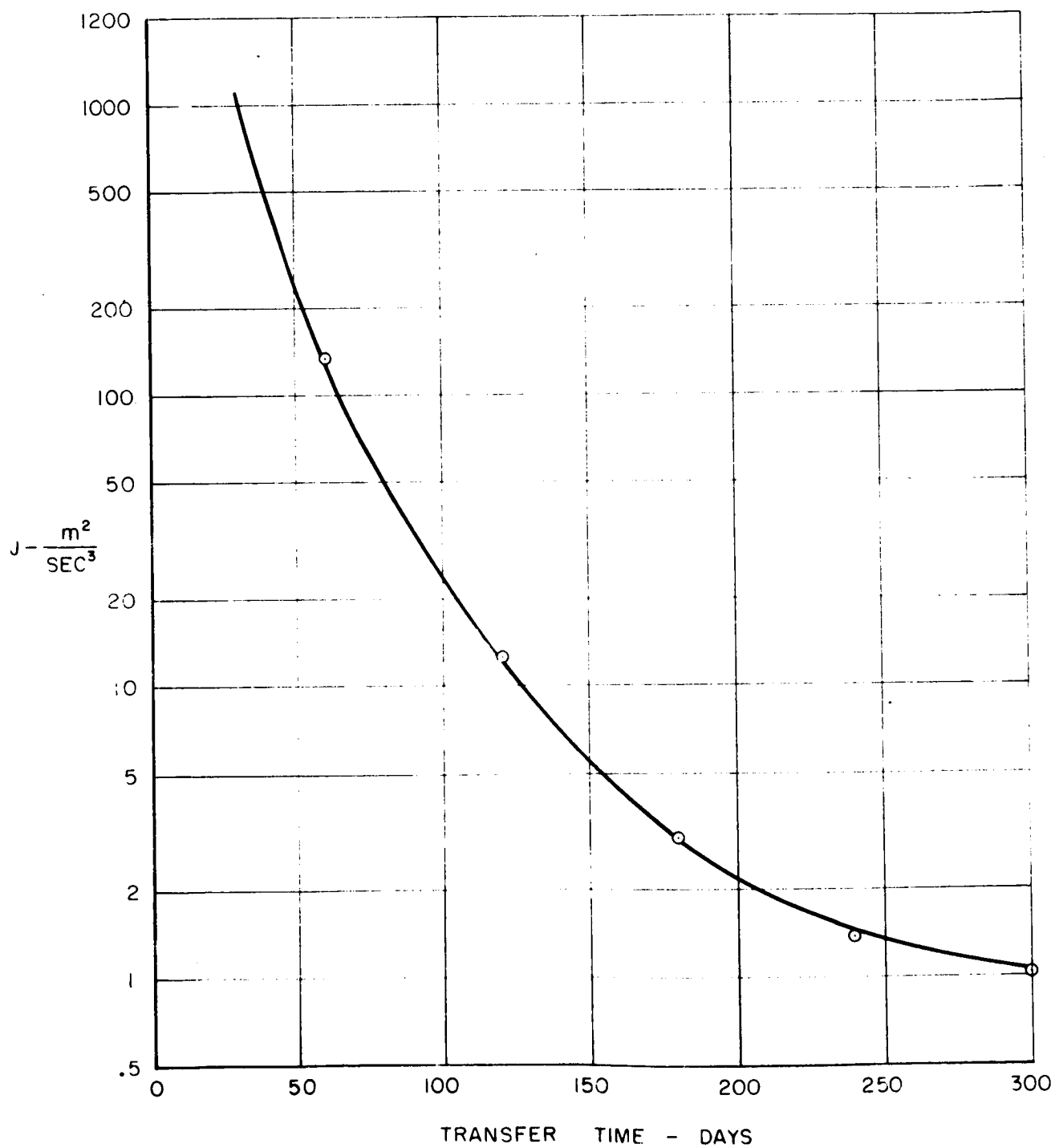




# OPTIMUM EARTH-VENUS TRANSFER

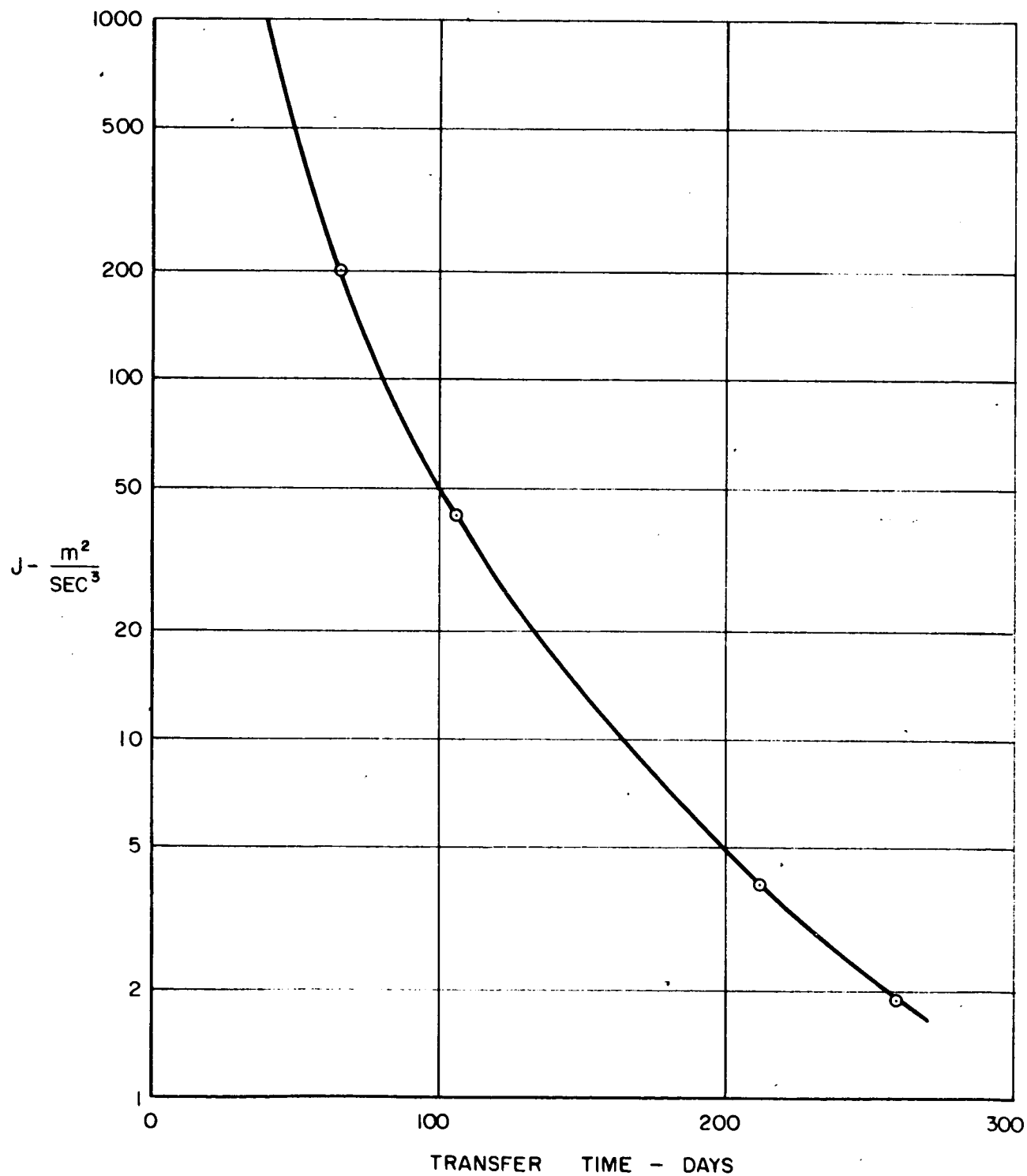
## UNINCLINED CIRCULAR TERMINAL ORBITS

○ - LINEARIZED ANALYSIS



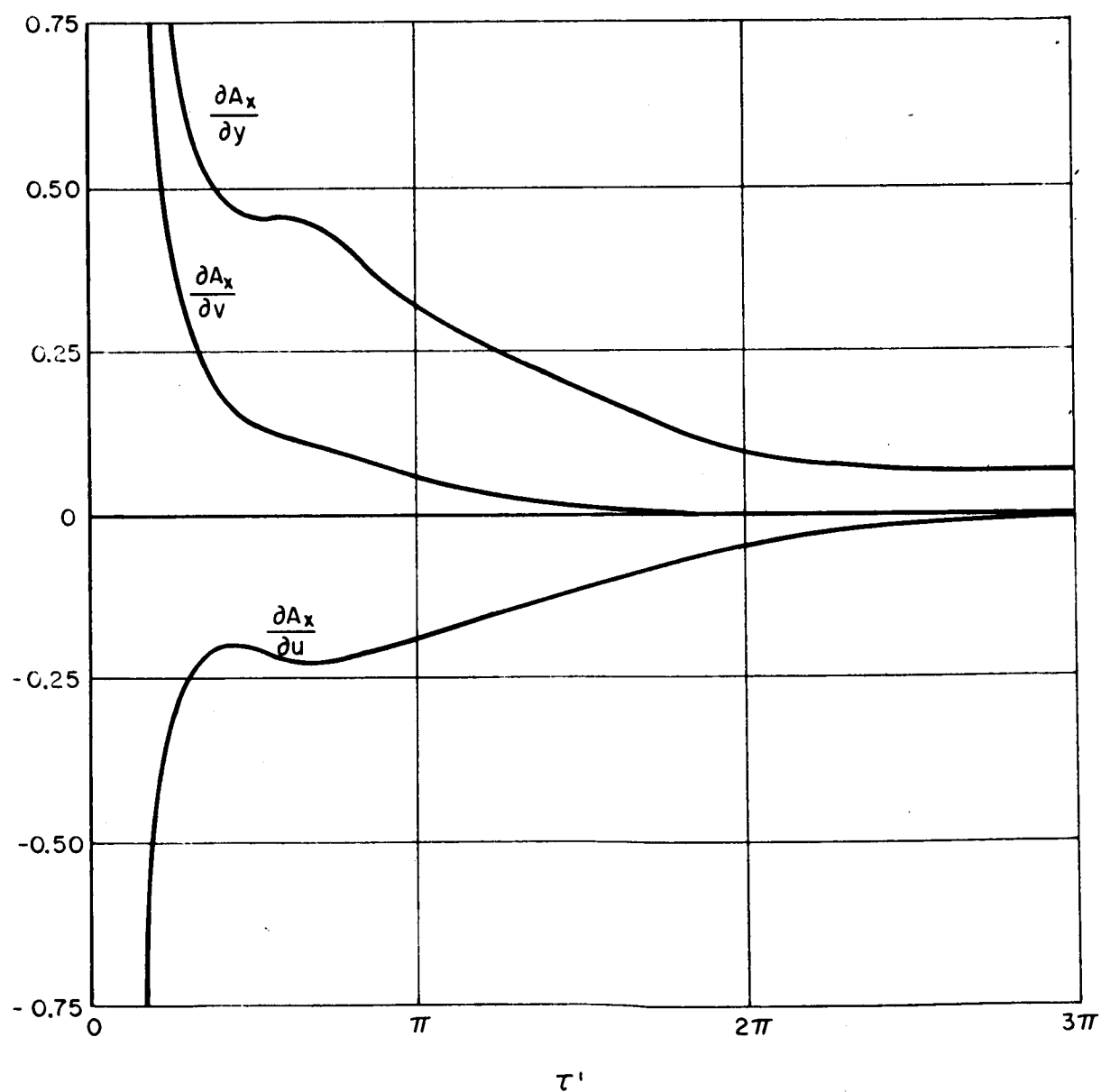
OPTIMUM EARTH - MARS TRANSFER  
UNINCLINED CIRCULAR TERMINAL ORBITS

○ - LINEARIZED ANALYSIS



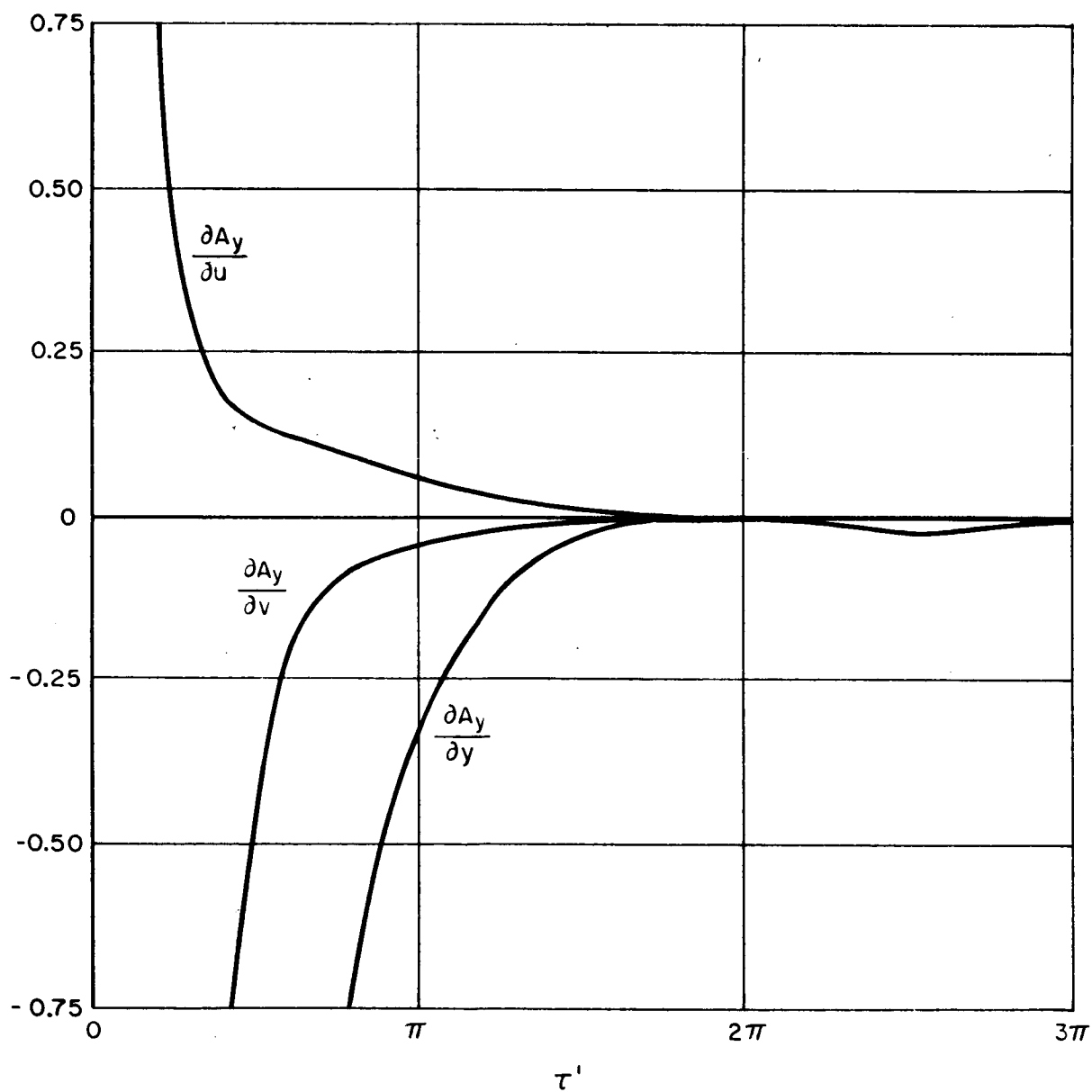
# GUIDANCE COEFFICIENTS FOR OPTIMUM CONTROL ORBIT TRANSFER

$$A_x = \frac{\partial A_x}{\partial y} y + \frac{\partial A_x}{\partial u} u + \frac{\partial A_x}{\partial v} v$$



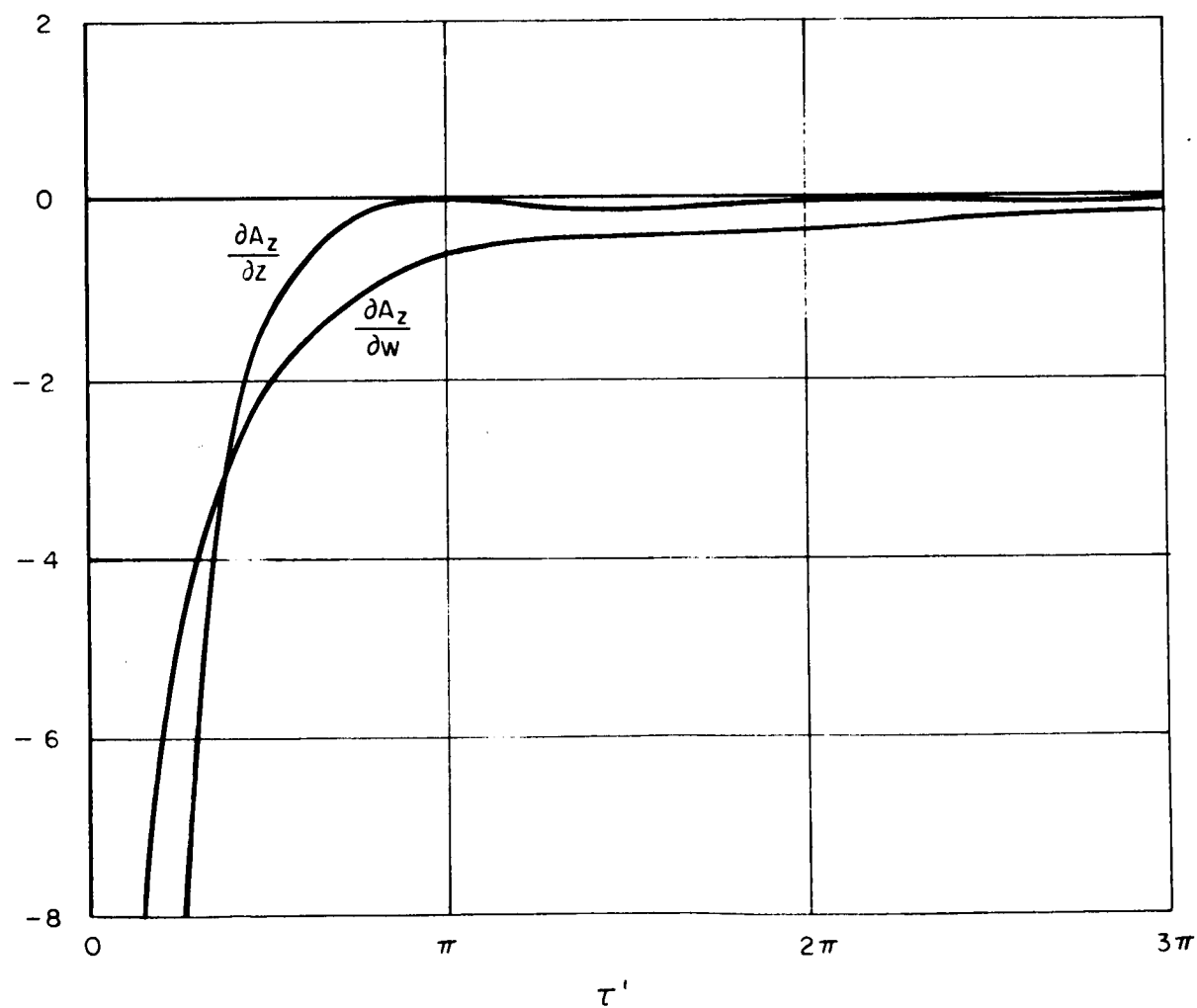
# GUIDANCE COEFFICIENTS FOR OPTIMUM CONTROL ORBIT TRANSFER

$$A_y = \frac{\partial A_y}{\partial y} y + \frac{\partial A_y}{\partial u} u + \frac{\partial A_y}{\partial v} v$$

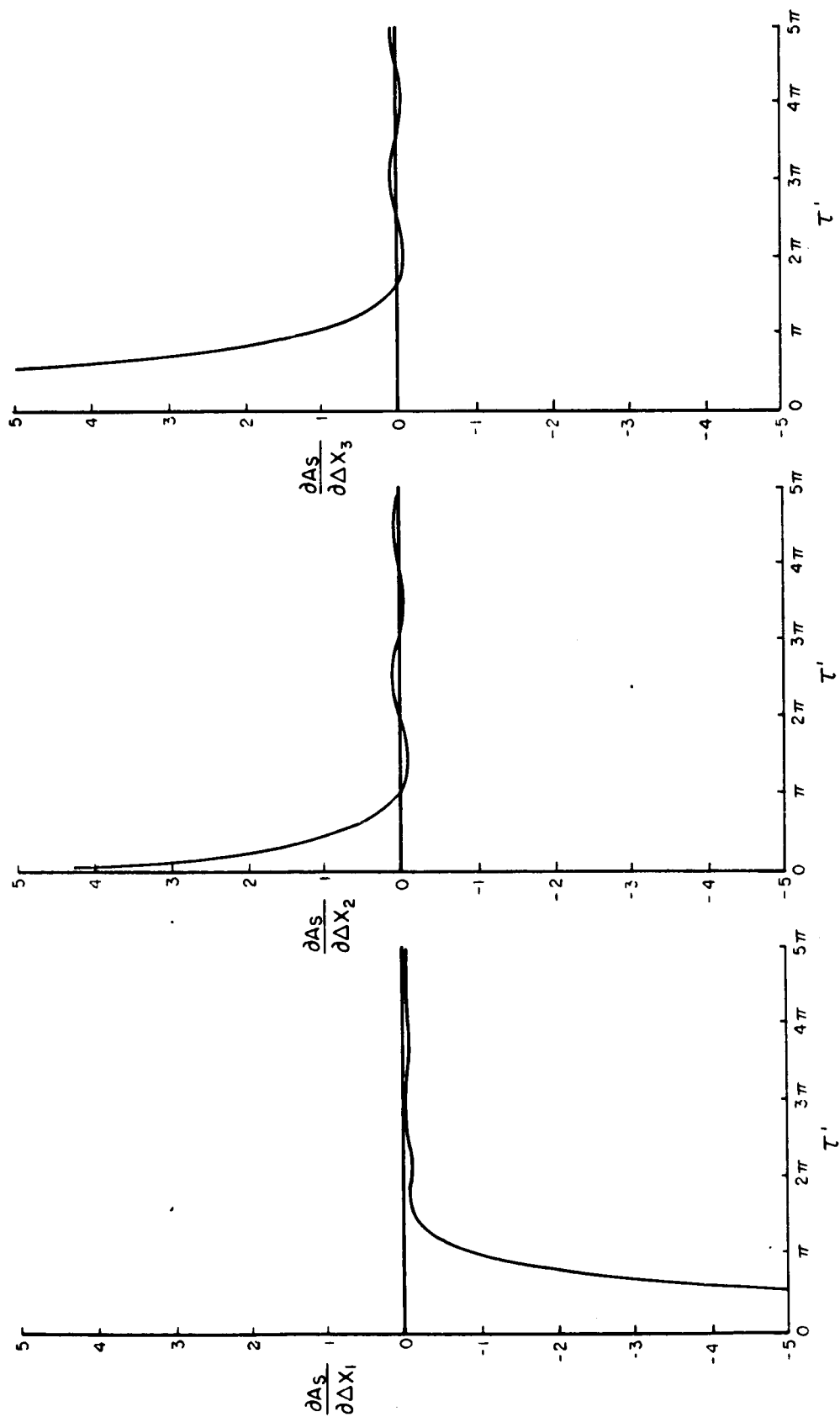


GUIDANCE COEFFICIENTS FOR OPTIMUM CONTROL  
ORBIT TRANSFER

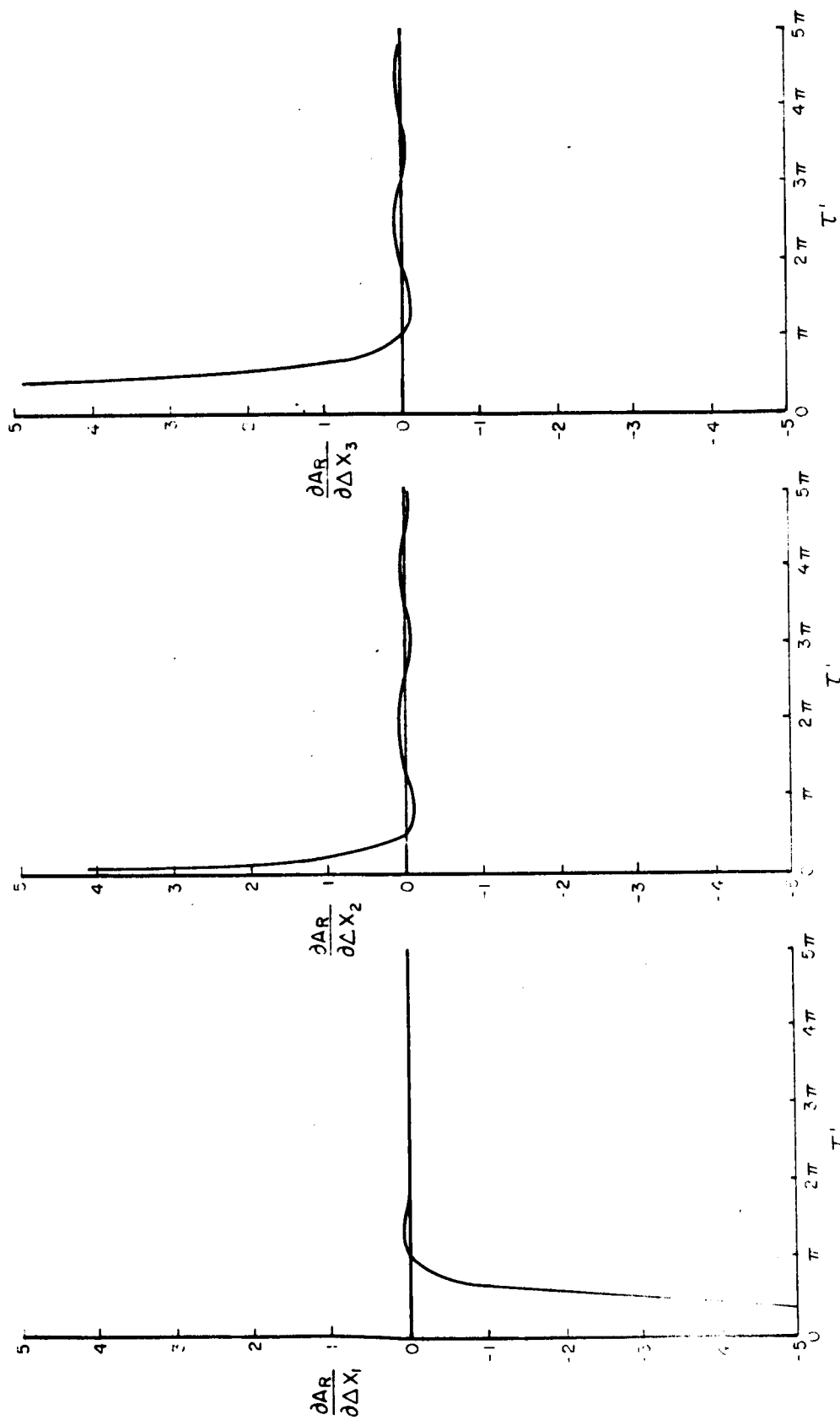
$$A_z = \frac{\partial A_z}{\partial w} w + \frac{\partial A_z}{\partial z} z$$



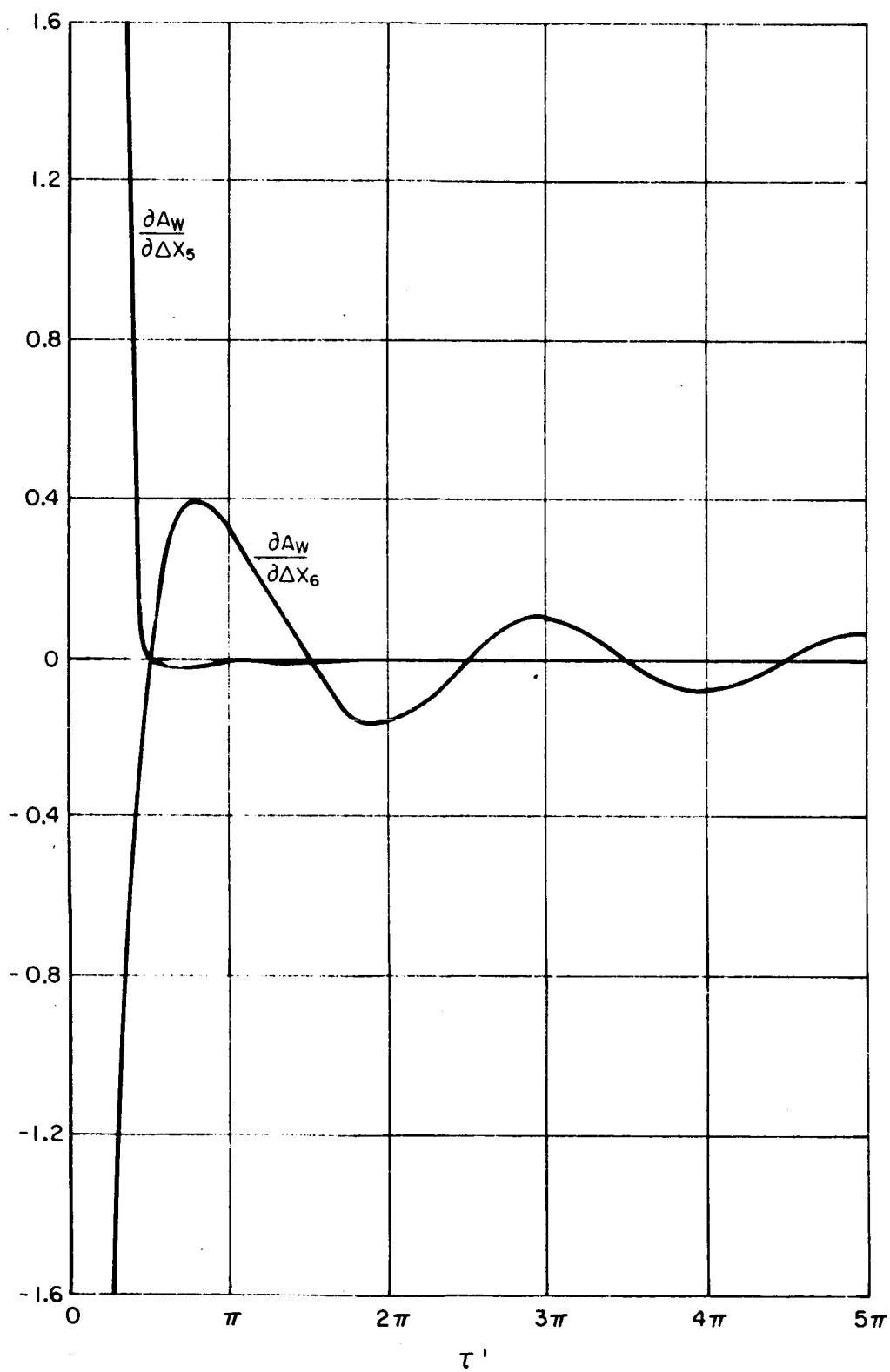
# GUIDANCE COEFFICIENTS FOR OPTIMUM CONTROL ORBIT TRANSFER



# GUIDANCE COEFFICIENTS FOR OPTIMUM CONTROL ORBIT TRANSFER

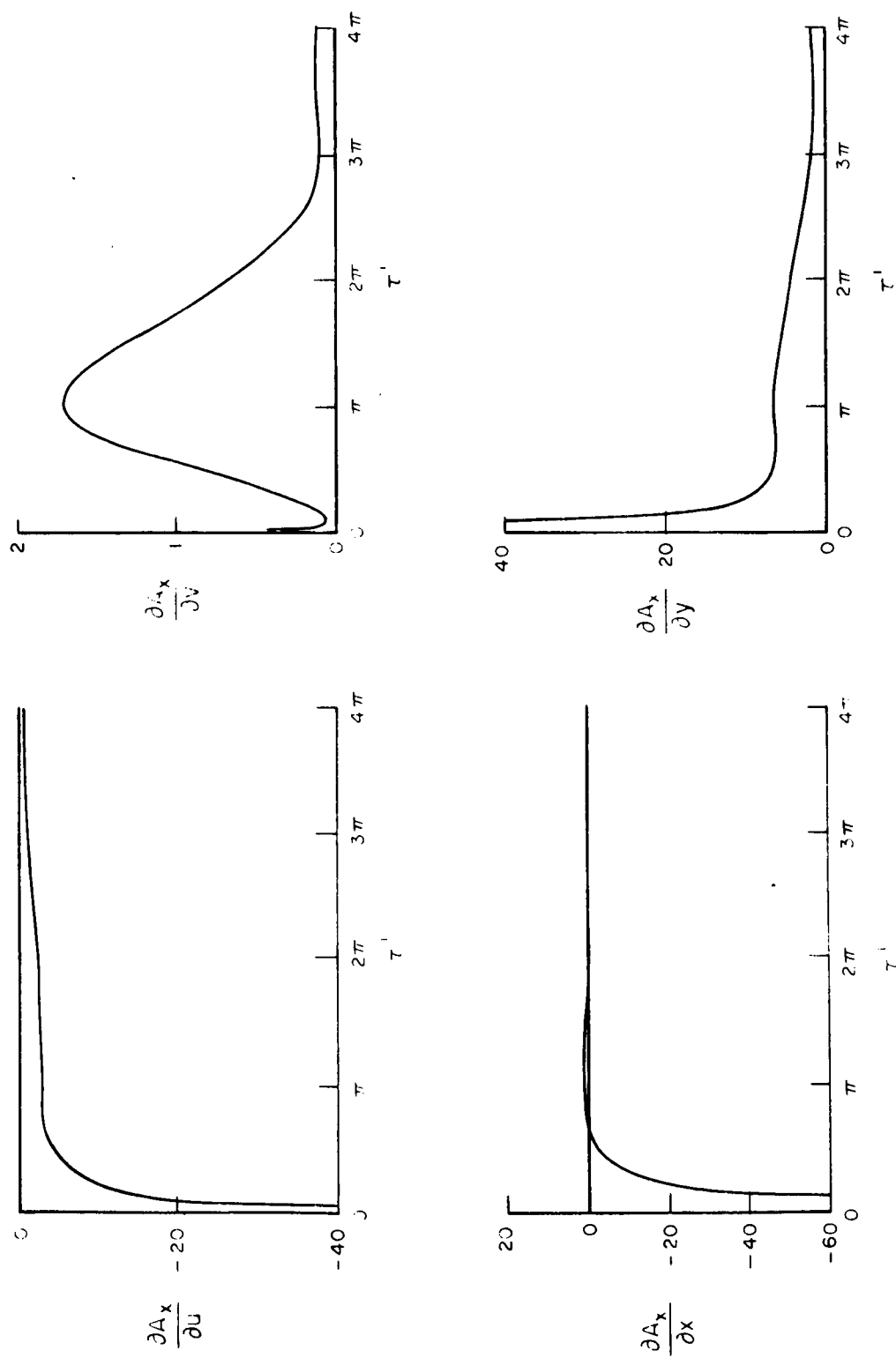


GUIDANCE COEFFICIENTS FOR OPTIMUM CONTROL  
ORBIT TRANSFER

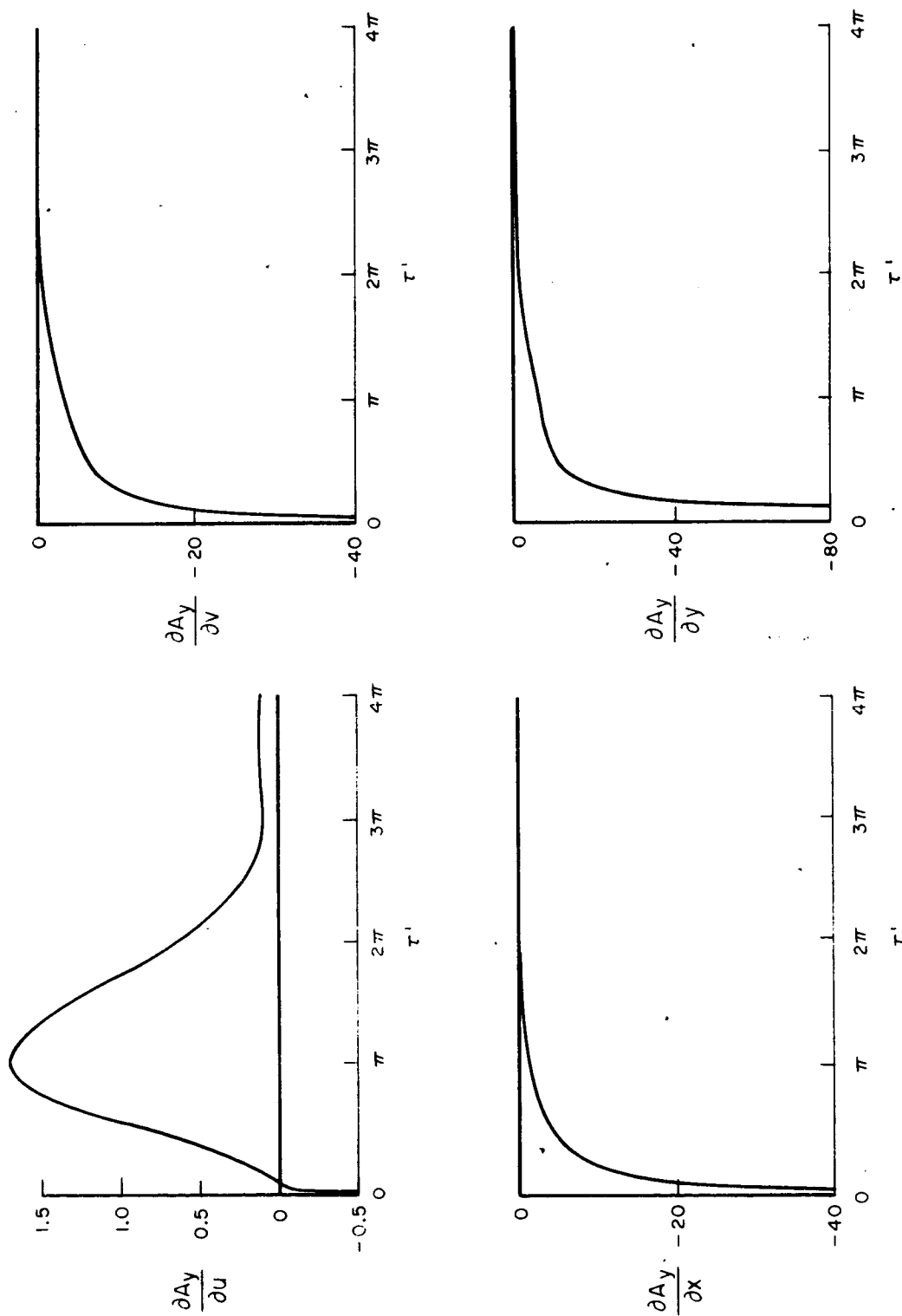




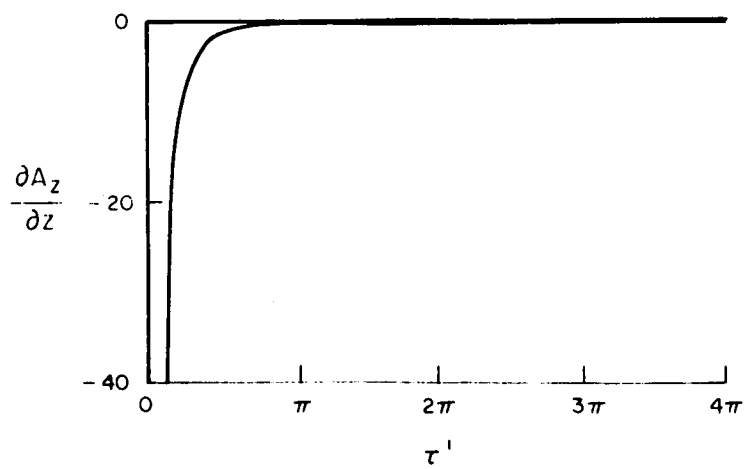
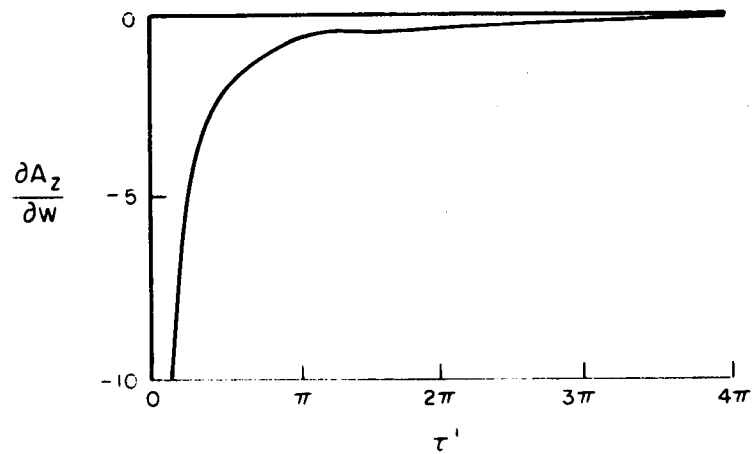
# GUIDANCE COEFFICIENTS FOR OPTIMUM CONTROL ORBITAL RENDEZVOUS



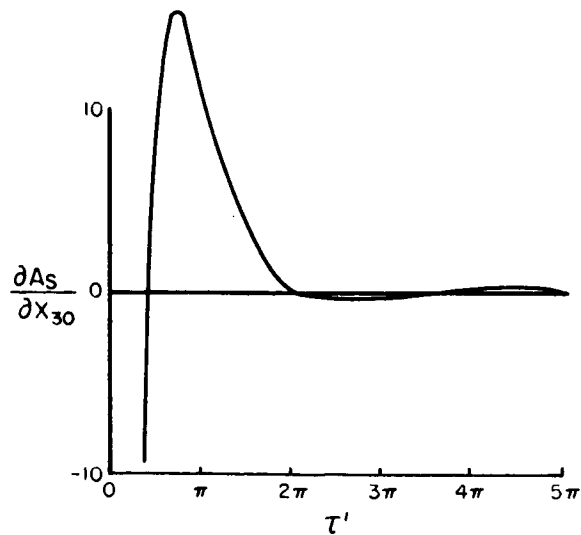
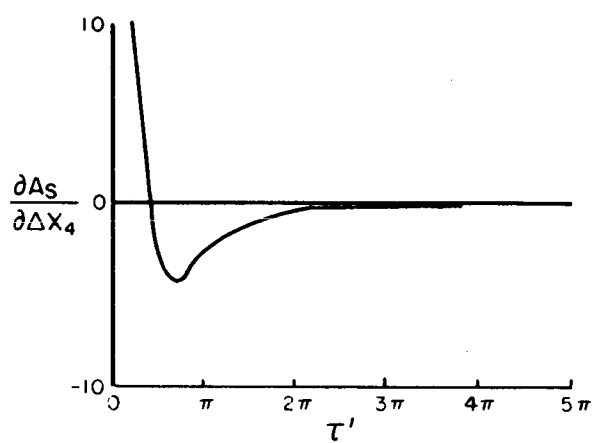
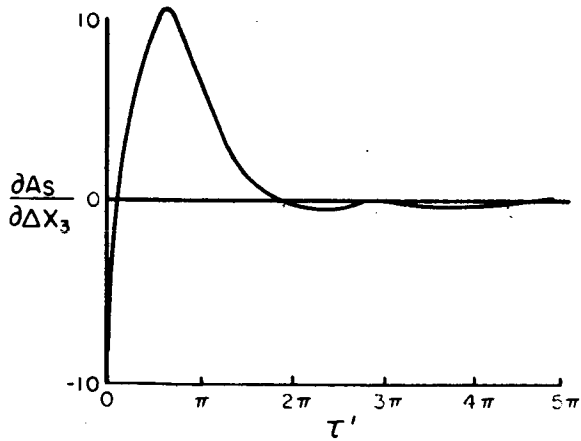
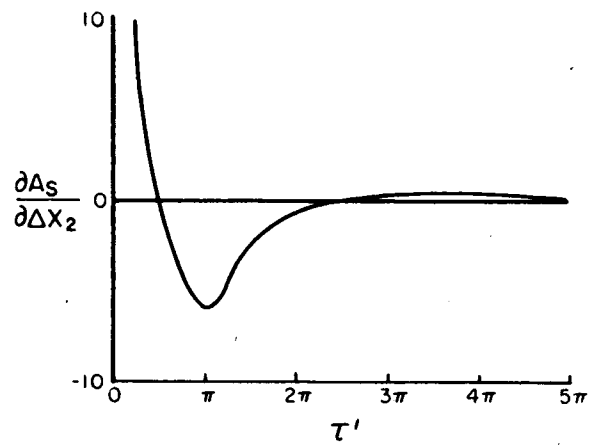
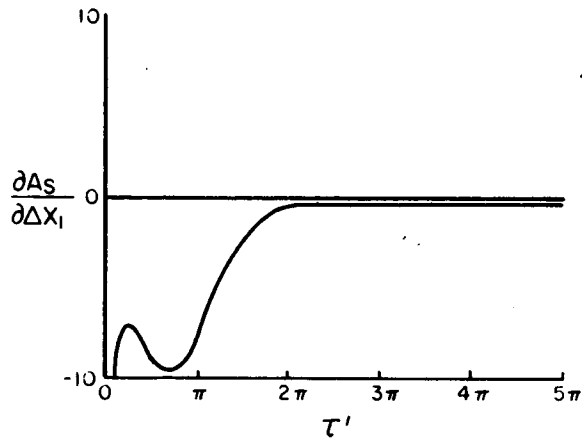
# GUIDANCE COEFFICIENTS FOR OPTIMUM CONTROL ORBITAL RENDEZVOUS



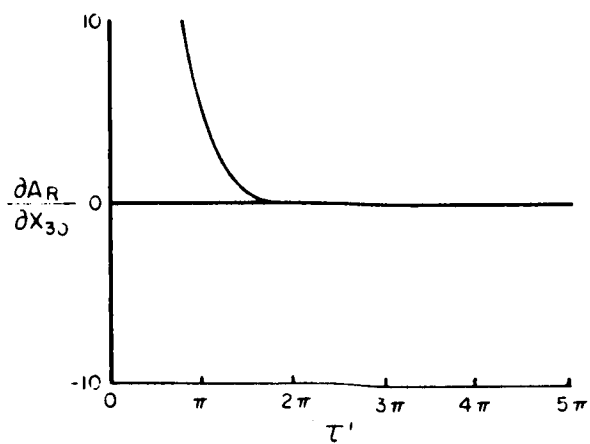
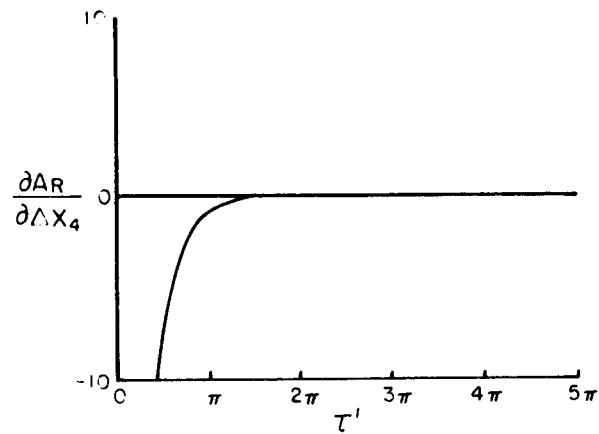
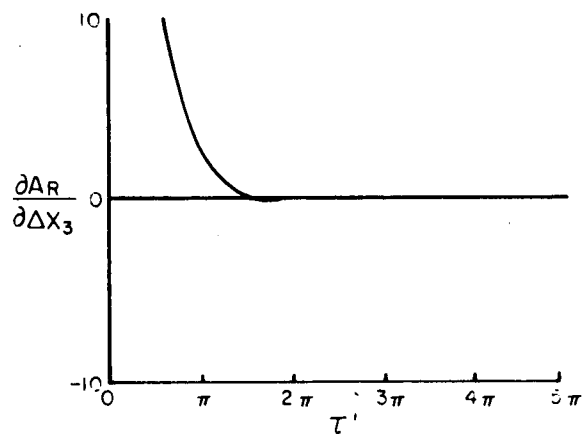
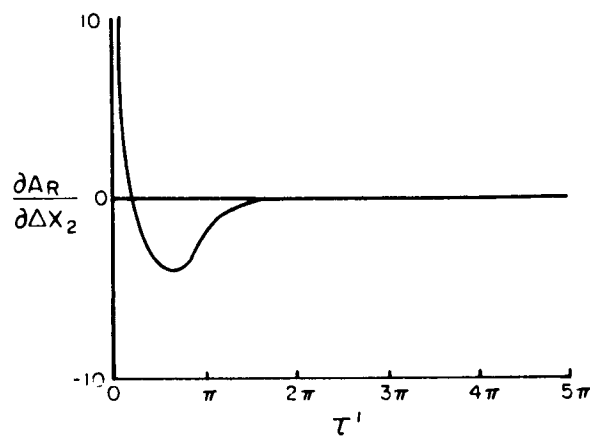
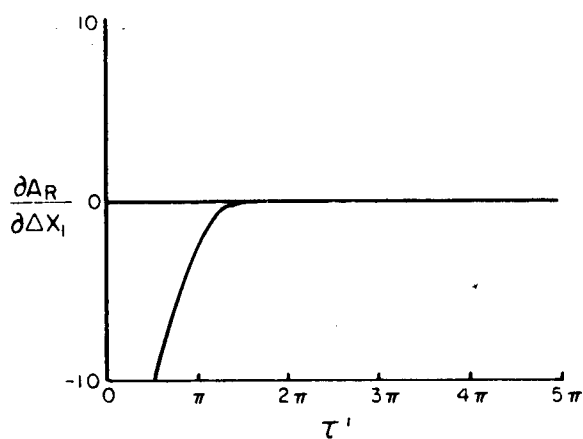
GUIDANCE COEFFICIENTS FOR OPTIMUM CONTROL  
ORBITAL RENDEZVOUS



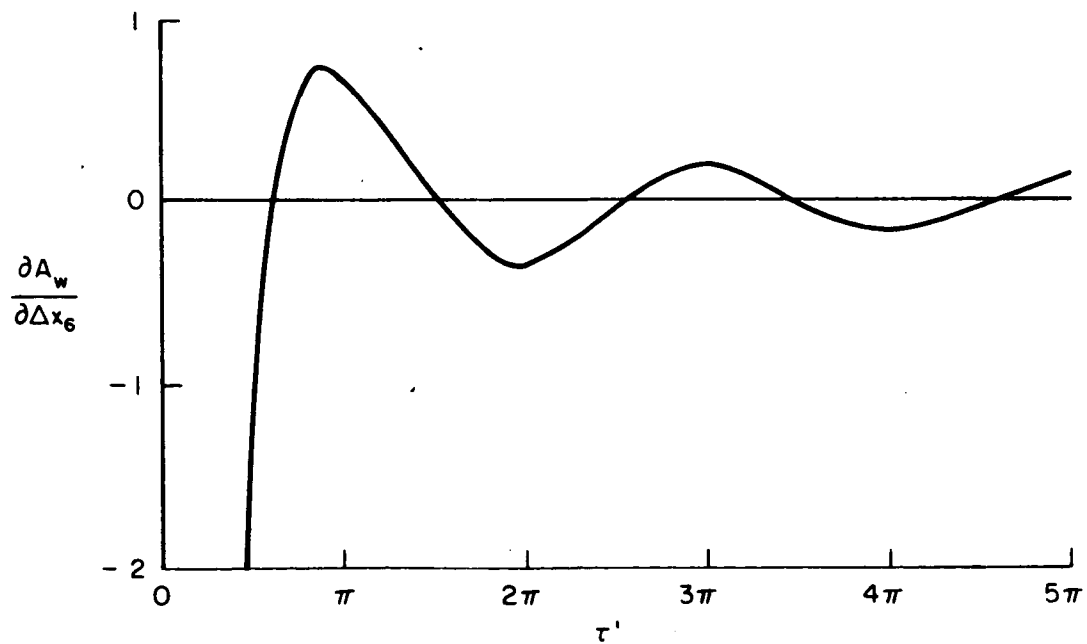
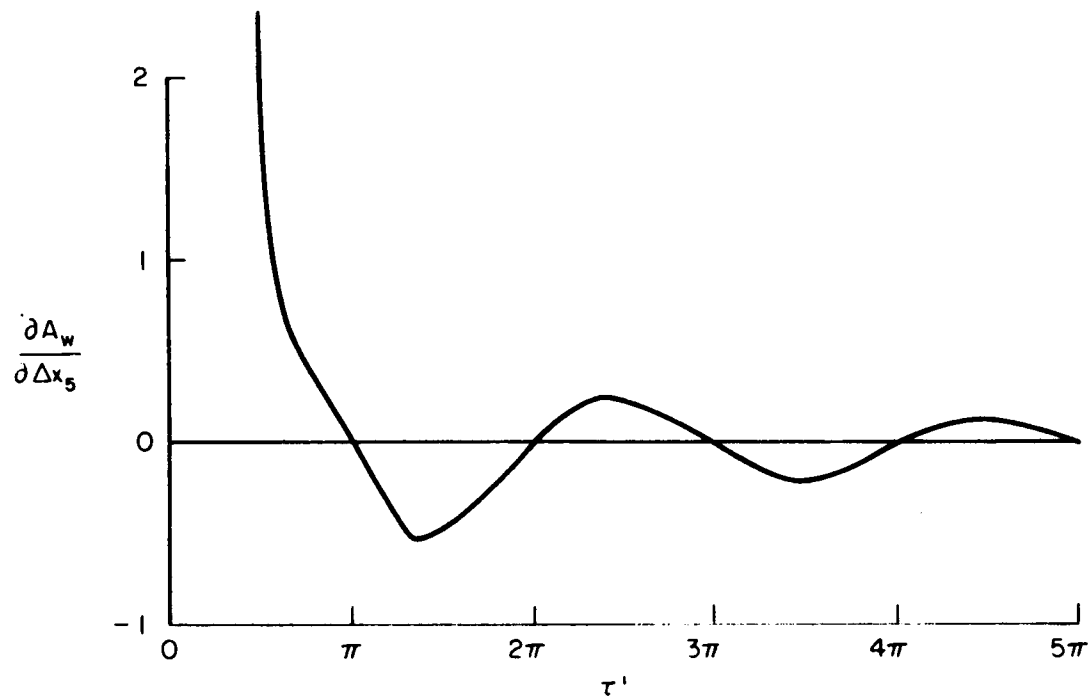
# GUIDANCE COEFFICIENTS FOR OPTIMUM CONTROL ORBITAL RENDEZVOUS



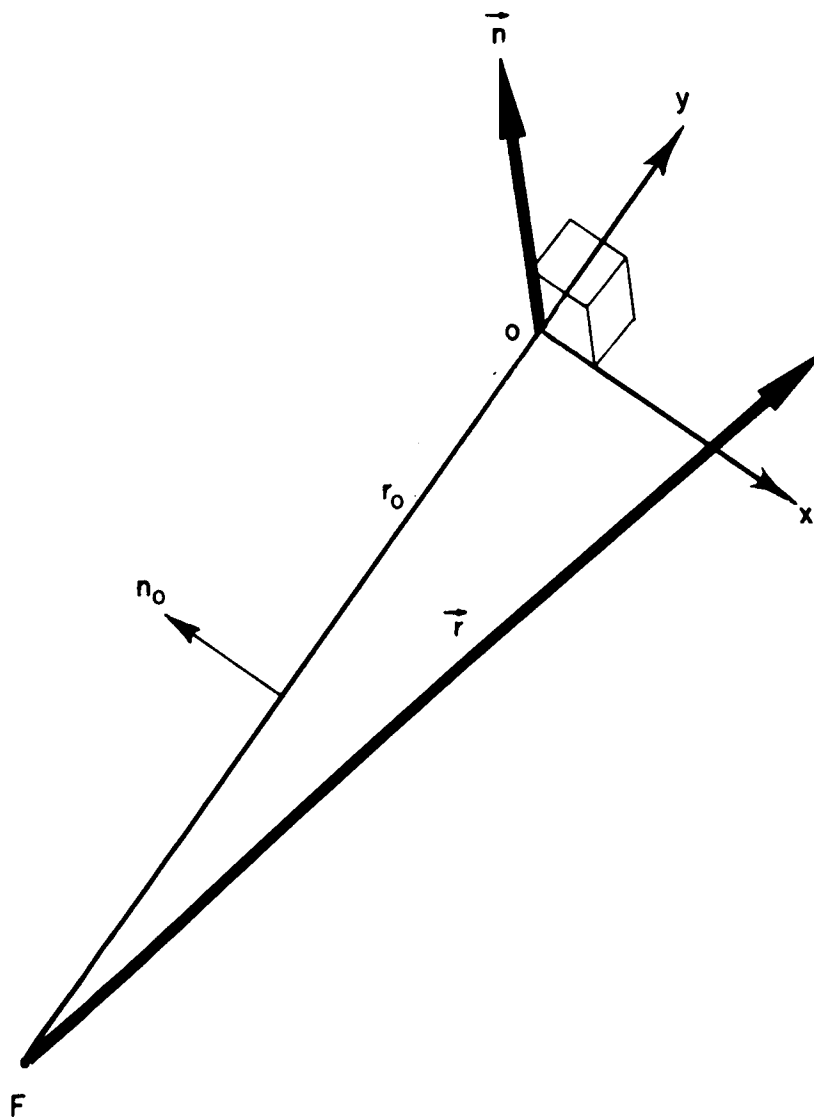
# GUIDANCE COEFFICIENTS FOR OPTIMUM CONTROL ORBITAL RENDEZVOUS



GUIDANCE COEFFICIENTS FOR OPTIMUM CONTROL  
ORBITAL RENDEZVOUS



RELATIONSHIP BETWEEN ROTATING AND NON-ROTATING  
COORDINATE SYSTEMS



N65 33057

HONEYWELL, INC.  
MILITARY PRODUCTS GROUP RESEARCH DEPARTMENT  
Minneapolis, Minnesota

AN APPROXIMATION TO LINEAR BOUNDED  
PHASE COORDINATE CONTROL PROBLEMS

By

E. B. Lee

Prepared for

George C. Marshall Space Flight Center  
National Aeronautics and Space Administration  
Huntsville, Alabama



# AN APPROXIMATION TO LINEAR BOUNDED PHASE COORDINATE CONTROL PROBLEMS\*

E. B. Lee

## 1. Introduction

In many control problems both restraints on the magnitudes of the control variables and various system variables may occur. Certain results [1,2,7] are available for the determination of optimal controllers for some classes of linear and nonlinear systems involving such restraints. These results take the form of necessary or sufficient conditions for optimal control but not both, and are therefore only a partial solution to even the theoretical problem, leaving much to be desired in the way of a practical solution. To use the necessary or sufficient conditions for synthesizing an optimal controller it is necessary to solve a two-point boundary value problem in terms of a number of free parameters and multipliers where the number of parameters is not even known as well as certain jump conditions [2,7]. A backing out procedure [9] is also available if one is interested in flooding the domain of controllability with responses and then keeping track (storing) of the corresponding control magnitude for each such point.

We here offer a procedure which has several advantages over the above schemes, but is only an approximate solution. Its main advantage is that no discontinuities will be encountered in the adjoint solution which determines the optimum controller and therefore the resulting two point boundary value problem may be more readily solved. The results provide both necessary and sufficient conditions, as well as existence,

\*Prepared under contract NASw-986 for the NASA.

for the approximate problem.

The analysis is limited to linear control processes as described by the differential system

$$1) \dot{x} = A(t)x + B(t)u(t).$$

The coefficient matrices  $A(t)$  and  $B(t)$  are composed of known continuous functions on the time interval  $[t_0, t_1]$ . The controller  $u(t)$  is to be chosen from a set  $\Omega: |u^j| \leq 1; j = 1, 2, \dots, m$ , so as to steer the response,  $x_u(t)$ , of 1) from an initial point  $x_0$  at time  $t_0$  to a prescribed compact target set  $\tilde{G} \subset R^n$  and it is required that  $x_u(t)$  remain within a given constraint set,  $\Lambda$ , during its entire response. Here  $R^n$  is the  $n$  dimensional real number space.

The problem of time optimal control, as considered in the next section, is to find a controller  $u(t)$  which steers  $x_u(t)$  from  $x_0$  to  $\tilde{G} \subset \Lambda$  in minimum time, that is, minimizes  $C(u) = t_1 - t_0$  with  $x(t_1) \in \tilde{G}$  and  $x_u(t) \in \Lambda, t_0 \leq t \leq t_1$ . Later, in section 4, we discuss other optimum control cost functionals.

There are certain difficulties involved when one directly solves for this optimum controller. We shall therefore be content with solving the following apparently simpler problem: Find that controller  $u(t)$  with graph in  $\Omega$  which steers  $x_u(t)$  from  $x_0$  at  $t_0$  to  $\tilde{G}$  at  $t_1$  with  $x_u^0(t_1) \leq \beta$  and  $t_1 - t_0$  a minimum.  $x_u^0(t)$  is defined below.

It is assumed that  $\Lambda$  is a closed convex set, (for convenience we could even let  $\Lambda = \{x | x'Hx \leq c\}$ , where  $H$  is a positive semi-definite matrix and  $c = \text{constant} > 0$ .) Let  $F(x)$  be a convex continuous differentiable function which is such that

$$\begin{aligned} F(x) &\neq 0 && \text{if } x \notin \Lambda \\ &= 0 && \text{if } x \in \Lambda \end{aligned}$$

Then define<sup>†</sup>

$$x_u^o(t_1) = \int_{t_0}^{t_1} F(x_u(t)) dt.$$

$x_u^o(t_1)$  essentially measures the excursions of the response  $x_u(t)$  to a controller  $u(t)$  outside of the region  $\Lambda$  during the time interval  $[t_0, t_1]$ . By keeping  $x_u^o(t_1)$  small the response  $x_u(t)$  is restricted to stay close to or within  $\Lambda$ . The above minimum time optimal control problem is approximately solved by finding a controller which steers  $\hat{x}_u(t) = (x_u^o(t), x_u(t))$  from  $(0, x_0)$  to  $G = \{x^o, x | x \in \tilde{G}, 0 \leq x^o \leq \beta\}$  in the minimum time interval  $t_1 - t_0$  if  $\beta > 0$  is sufficiently small.

In the next section we give necessary and sufficient conditions for this approximation problem using the time optimal criterion. Section 3 contains an example and section 4 is a discussion of the approximation problem for other cost functionals.

## 2. The necessary and sufficient conditions for the approximate linear time optimal problems

We augment the system  $f$  by considering the equation system

---

<sup>†</sup>There is, of course, some question as to whether such a function  $F(x)$  exists for an arbitrary convex set  $\Lambda$  contained in  $R^n$ . We now cite an example which shows that there are such functions in a number of interesting cases. Suppose  $\Lambda = \{x^1, x^2, \dots, x^n | |x^2| \leq 1\}$ . Then pick  $F(x)$

$$\begin{aligned} &= 1/2(x^2 - 1)^2 && \text{if } x^2 \geq 1 \\ &= 0 && \text{if } |x^2| \leq 1 \\ &= 1/2(x^2 + 1)^2 && \text{if } x^2 \leq -1 \end{aligned}$$

Thus if only one coordinate (or a linear combination) is restricted the problem is easily handled as in the example, where  $F(x)$  is continuous and has continuous partial derivatives. Other  $\Lambda$ 's can be approximately handled as in the example.

$$\hat{f}) \quad \dot{x}^0 = F(x)$$

$$\dot{x} = A(t)x + B(t) u(t)$$

obtained from  $\hat{f}$ ) by adding the equation for  $\dot{x}^0$  with  $x^0(t_0) = 0$ . Here  $A(t)$ ,  $B(t)$  are bounded and continuous on  $[t_0, t_1]$  and  $F(x)$  is a convex function with  $F(x) = 0$  for  $x \in \Lambda$ .  $\frac{\partial F}{\partial x}(x)$  is assumed to exist and be continuous everywhere.

The set of attainability  $\hat{K}(t_1) \subset R^{n+1}$  is the collection of end points  $\hat{x}_u(t_1)$  of responses  $\hat{x}_u(t) = (x_u^0(t), x_u(t))$  of  $\hat{f}$  which initiate at  $(0, x_0)$  at time  $t_0$  corresponding to all (Lebesgue) measurable controllers  $u(t)$  which are such that  $|u^j(t)| \leq 1$  on  $[t_0, t_1]$ , for  $j = 1, 2, \dots, m$ . (Such controllers are referred to as admissible controllers.)

In the following theorems we establish various properties for  $\hat{K}(t_1)$  and  $\partial\hat{K}(t_1)$  as required in synthesizing optimal controllers.

Theorem 1 Consider the above system  $\hat{f}$ ) with initial point  $\hat{x}_0$ , restraint set  $\Omega$ , and set of attainability  $\hat{K}(t_1)$ .

Then  $\hat{K}(t_1)$  is a nonempty compact subset of  $R^{n+1}$  in variables  $(x^0, x)$  with convex lower surface (as defined below) for each  $t_0 \leq t_1 < \infty$ .

Proof  $\hat{K}(t_1)$  is nonempty since any measurable controller  $u(t) \subset \Omega$  gives rise to an end point  $\hat{x}_u(t_1) \in \hat{K}(t_1)$ .  $\hat{K}(t_1)$  is compact because the system  $\hat{f}$ ) satisfies the hypothesis of the existence theorems of references 6, and 8.

The lower surface of  $\hat{K}(t)$  is where exterior normal  $n+1$  vectors  $\hat{\eta}$  to  $\hat{K}(t)$  at points of  $\partial\hat{K}(t)$  have their first component  $\eta_0 \leq 0$ . We now show that if  $\hat{x}_1$  and  $\hat{x}_2$  are points of  $\hat{K}(t_1)$  then the point  $\hat{y} = \lambda\hat{x}_1 + (1-\lambda)\hat{x}_2 = (y^0, y)$ ,  $0 \leq \lambda \leq 1$ , is such that

$$y = x_{\bar{u}}(t_1)$$

and

$$y^0 \geq x_{\bar{u}}^2(t_1),$$

where  $\bar{u}(t) = \lambda u_1(t) + (1-\lambda) u_2(t)$  and  $u_1(t)$  and  $u_2(t)$  are such that  $\hat{x}_{u_1}(t_1) = \hat{x}_1$  and  $\hat{x}_{u_2}(t_1) = \hat{x}_2$ . The convexity of the lower surface of  $\hat{K}(t_1)$  then follows because in order for it to be nonconvex it is necessary that there exist two points  $\hat{x}_1, \hat{x}_2$  on this lower boundary, with the property that the point  $\lambda \hat{x}_1 + (1-\lambda) \hat{x}_2$  is below the set  $\hat{K}(t_1)$  for some  $0 < \lambda < 1$ , which will then be impossible.

With  $\bar{u}(t) = \lambda u_1(t) + (1-\lambda) u_2(t)$  we find that

$$\begin{aligned} x_{\bar{u}}(t_1) &= \Phi(t_1)x_0 + \int_{t_0}^{t_1} \Phi(t_1)\Phi^{-1}(s)B(s)\bar{u}(s)ds \\ &= \lambda \left[ \Phi(t_1)x_0 + \int_{t_0}^{t_1} \Phi(t_1)\Phi^{-1}(s)B(s)u_1(s)ds \right] \\ &\quad + (1-\lambda) \left[ \Phi(t_1)x_0 + \int_{t_0}^{t_1} \Phi(t_1)\Phi^{-1}(s)B(s)u_2(s)ds \right] \\ &= \lambda x_{u_1}(t_1) + (1-\lambda) x_{u_2}(t_1) = \\ &= \lambda x_1 + (1-\lambda) x_2 = y \end{aligned}$$

where  $\Phi(t)$  is the fundamental solution matrix of  $\dot{x} = f(x, u)$  with  $\Phi(t_0) = I$ . We also calculate

$$x_{\bar{u}}^2(t_1) = \int_{t_0}^{t_1} F(x_{\bar{u}}(t))dt$$

and  $\lambda x_{u_1}^o(t_1) + (1-\lambda) x_{u_2}^o(t_1)$  for comparison. Since  $F(x)$  is a convex function of  $\hat{x}$  it follows that for  $0 \leq \lambda \leq 1$ ,

$$F(x_{\bar{u}}(t)) = F(\lambda x_{u_1}(t) + (1-\lambda)x_{u_2}(t)) \leq \lambda F(x_{u_1}(t)) + (1-\lambda) F(x_{u_2}(t))$$

and so

$$\begin{aligned} x_{\bar{u}}^o(t_1) &= \int_{t_0}^{t_1} F(x_{\bar{u}}(t)) dt = \int_{t_0}^{t_1} F(\lambda x_{u_1}(t) + (1-\lambda) x_{u_2}(t)) dt \\ &\leq \lambda \int_{t_0}^{t_1} F(x_{u_1}(t)) dt + \int_{t_0}^{t_1} (1-\lambda) F(x_{u_2}(t)) dt = y^o. \end{aligned}$$

Q.E.D.

We will now consider those controllers  $u(t)$  on  $[t_0, t_1]$  which steer  $\hat{x}_u(t)$  from  $\hat{x}_0$  at  $t_0$  to points  $\hat{x}_1$  contained in the lower boundary of  $\hat{K}(t_1)$  (written  $\partial\hat{K}^-(t_1)$ ). Such controllers will be called extremal and they will play a significant part in the selection of optimal controllers.

Let  $u(t) \in \Omega$  on  $t_0 \leq t \leq t_1$  be an admissible controller for the convex control process

$$\hat{\dot{x}} = F(x)$$

$$\dot{x} = A(t) x + B(t)u(t)$$

with initial point  $\hat{x}_0 = (0, x_0)$  at  $t_0$ . If the corresponding response  $\hat{x}_u(t)$  has an end point  $\hat{x}(t_1) \in \partial\hat{K}^-(t_1)$ , then  $u(t)$  is called an extremal control and  $\hat{x}_u(t)$  an extremal response on  $[t_0, t_1]$ .

so that

$$\eta(t)B(t)u(t) = \underset{u \in \Omega}{\text{Max}} \{ \eta(t)B(t)u \} \text{ almost always on } [t_0, t_1].$$

[Proof: Assume  $u(t)$  on  $[t_0, t_1]$  is extremal and so steers  $\bar{x}(t)$  from  $(0, x_0)$  at  $t_0$  to  $\hat{x}_1 \in \partial \hat{K}^-(t_1)$  at  $t_1$ . Choose  $\hat{\eta}(t_1) = (\eta_0, \eta(t_1))$  to be a nonzero vector normal to  $\pi$  directed into the halfspace defined by  $\pi$  which does not meet  $\hat{K}(t_1)$ . Note  $\eta_0 < 0$ . Then let  $\hat{\eta}(t)$  with  $\hat{\eta}(t_1)$  as above be the response of the adjoint equation corresponding to the controller  $u(t)$ .

The controller  $^+ \bar{u}(t) = \text{sgn}\{\eta(t)B(t)\}$  defined for  $t \in [t_0, t_1]$  is admissible and

$$\eta(t)B(t)\bar{u}(t) = \underset{u \in \Omega}{\text{Max}} \{ \eta(t)B(t)u \}$$

on  $[t_0, t_1]$ .

Let  $\tau_\epsilon$  be an interval of total length  $\epsilon > 0$  contained in  $\mathcal{J} = [t_0, t_1]$  whereon

$$\delta + \eta(t)B(t)u(t) < \underset{u \in \Omega}{\text{Max}} \{ \eta(t)B(t)u \} \text{ for some } \delta > 0.$$

For given  $\delta > 0$  consider the modified controller

$$\begin{aligned} u_\epsilon(t) &= u(t) \text{ on } \mathcal{J} - \tau_\epsilon \\ &= \bar{u}(t) \text{ on } \tau_\epsilon, \end{aligned}$$

---


$$\begin{aligned} + \text{sgn} \{ \} &= -1 \quad \text{if } \{ \} < 0 \\ &= 0 \quad \text{if } \{ \} = 0 \\ &= +1 \quad \text{if } \{ \} > 0 \end{aligned}$$

The adjoint response  $\hat{\eta}(t) = (\eta_0(t), \eta(t))$  corresponding to a controller  $u(t)$  is a row  $n+1$  vector satisfying the differential system

$$\dot{\eta} = -\eta A(t) - \eta_0 \frac{\partial F'}{\partial x} (x_u(t))$$

$$\eta_0 = \text{constant} \lesseqgtr 0.$$

where  $x_u(t)$  is the response of  $\mathcal{J}$  corresponding to the controller  $u(t)$ . Define  $u(t)$  on  $[t_0, t_1]$  to be a maximal controller in case there exists a nonvanishing adjoint response  $\hat{\eta}(t)$ ,  $\eta_0 \leq 0$ , so that  $\eta(t)B(t)u(t) = \text{Max}_{u \in \Omega} \{\eta(t)B(t)u\}$  a.e. on  $[t_0, t_1]$ .

In the following theorem 2 it is shown that extremal and maximal controllers are the same.

Theorem 2 Consider the convex control process†

$$\hat{\mathcal{J}}) \quad \dot{x}^0 = F(x)$$

$$\dot{x} = A(t)x + B(t)u(t)$$

with initial point  $\hat{x}_0 = (0, x_0)$  at time  $t_0$ . An admissible controller  $u(t) \in \Omega$  on  $[t_0, t_1]$  is extremal for  $\hat{\mathcal{J}}$  if and only if it is a maximal controller, that is, if and only if there is a nonvanishing adjoint response  $\hat{\eta}(t)$  of

$$\dot{\eta} = -\eta A(t) - \eta_0 \frac{\partial F'}{\partial x} (x_u(t))$$

$$\eta_0 = \text{constant} \lesseqgtr 0$$

---

†The necessary portion of this theorem follows from L. S. Pontryagin's Maximum Principle (7). For completeness the simple arguments to establish the necessary part are presented.



and calculate

$$\frac{d\hat{\eta}(t)\hat{x}_\epsilon}{dt} = \dot{\hat{\eta}}\hat{x}_\epsilon + \hat{\eta}\dot{\hat{x}}_\epsilon$$

and

$\frac{d\hat{\eta}(t)\hat{x}}{dt} = \dot{\hat{\eta}}\hat{x} + \hat{\eta}\dot{\hat{x}}$ , where  $\hat{x}_\epsilon$  refers to a response of  $\hat{f}$  corresponding to the modified controller  $u_\epsilon(t)$ .

Integration from  $t_0$  to  $t_1$  yields

$$\begin{aligned} \hat{\eta}(t_1)\hat{x}_\epsilon(t_1) - \hat{\eta}(t_0)\hat{x}_\epsilon(t_0) &= \int_{t_0}^{t_1} \left[ -\eta A(t) + \frac{\partial F}{\partial x}(x(t)) \right] x_\epsilon(t) \\ &+ \int_{t_0}^{t_1} \eta(t) \left[ A(t)x_\epsilon(t) + B(t)u(t) \right] - F(x_\epsilon(t)) dt \end{aligned}$$

and

$$\begin{aligned} \hat{\eta}(t_1)\hat{x}(t_1) - \hat{\eta}(t_0)\hat{x}(t_0) &= \int_{t_0}^{t_1} \left\{ \left[ -\eta A(t) + \frac{\partial F}{\partial x}(x(t)) \right] x(t) \right. \\ &+ \left. \eta(t) \left[ A(t)x(t) + B(t)u(t) \right] - F(\dot{x}(t)) \right\} dt \text{ for } \eta_0 = -1. \end{aligned}$$

Combining terms and using the assumed continuity for  $F$  and  $\frac{\partial F}{\partial x}$  we easily find that

$\hat{\eta}(t_1)\hat{x}_\epsilon(t_1) - \hat{\eta}(t_1)\hat{x}(t_1) \geq \delta \epsilon + o(\epsilon)$  for  $\epsilon$  sufficiently small where  $o(\epsilon)$  corresponds to terms of higher than first order in  $\epsilon$ , and therefore for  $\epsilon$  sufficiently small

$\hat{\eta}(t_1)x_\epsilon(t_1) - \hat{\eta}(t_1)\hat{x}(t_1) > 0$ , contradicting the construction of  $\hat{\eta}(t_1)$  as the outward normal to  $K(t_1)$  at  $\hat{x}_1$ .

Hence there exists no such interval  $\tau_\epsilon$ , so

$$\eta(t)B(t)u(t) = \text{Max}_{u \in \Omega} \eta(t)B(t)u \text{ almost everywhere on } \mathcal{J}.$$

Conversely, assume that  $u(t)$  and corresponding response  $\hat{\eta}(t) \neq 0$  are such that

$$\eta(t)B(t)u(t) = \text{Max}_{u \in \Omega} \eta(t)Bu$$

a.e. on  $\mathcal{J}$  with  $\eta_0 \geq 0$ . Let  $\bar{u}(t)$  be any controller in  $\Omega$  with corresponding response  $x_{\bar{u}}(t)$ . If we calculate

$$\frac{d\hat{\eta}\hat{x}_u}{dt} \text{ and } \frac{d\hat{\eta}\hat{x}_{\bar{u}}}{dt} \text{ as above,}$$

and then integrate from  $t_0$  to  $t_1$  using the assumed convexity of  $F(x)$  we find that

$$\hat{\eta}(t_1) \hat{x}_u(t_1) \geq \hat{\eta}(t_1) \hat{x}_{\bar{u}}(t_1) = \hat{\eta}(t_1) \hat{w}$$

where  $\hat{w}$  is any point of  $\hat{K}(t_1)$ . Since  $|\hat{\eta}(t_1)| \neq 0$ , and  $\eta_0 \leq 0$ , the above inequality implies that  $\hat{x}_u(t_1)$  is contained in the lower boundary of the compact set  $\hat{K}(t_1)$  with convex lower boundary and hence  $u(t)$  is extremal QED.

Theorem 2 indicates that to stay at a lower boundary point we must continuously steer maximally in the direction of the vector  $\hat{\eta}(t)$ . This remark is summarized as a corollary. Corollary 2.1 Let  $u(t)$  on  $[t_0, t_1]$  be an extremal controller for  $\hat{\mathcal{I}}$ , with corresponding response  $\hat{x}_u(t)$  and adjoint response  $\hat{\eta}(t)$  so that,

$$\eta(t)B(t)u(t) = \text{Max}_{u \in \Omega} \eta(t)B(t)u$$

a.e. on  $[t_0, t_1]$ . Then on each subinterval  $[t_0, \tau] \subset [t_0, t_1]$ ,  $u(t)$  is also an extremal controller with  $\hat{x}_u(\tau) \in \partial \hat{K}(\tau)$ .

Moreover  $\hat{\eta}(\tau)$  is an exterior normal to  $\hat{K}(\tau)$  at  $\hat{x}(\tau)$ .

Proof Replace  $t_1$  by  $\tau$  in the proof of theorem 2 to obtain that

$$\hat{\eta}(\tau) \hat{x}_u(\tau) \geq \hat{\eta}(\tau) \hat{x}_w(\tau) = \hat{\eta}(\tau) \hat{w}(\tau)$$

for all  $\hat{w}(\tau)$  in  $\hat{K}(\tau)$ . From this inequality the conclusion of the corollary can be drawn.

We next show that the set of attainability  $\hat{K}(t_1)$  depends continuously on the parameter  $t_1$ .

Define the distance between a point  $p$  and a compact set  $G_1 \subset \mathbb{R}^n$  to be

$$d(p, G_1) = \min_{g \in G_1} |p - g|$$

and define the distance between two compact sets  $G_1$ , and  $G_2 \subset \mathbb{R}^n$  to be

$$d(G_2, G_1) = \max \left\{ \max_{p_1 \in G_1} d(p_1, G_2), \max_{p_2 \in G_2} d(p_2, G_1) \right\}. \text{ Here}$$

$$|p| = \sum_{i=1}^n |p^i|.$$

The set  $\hat{K}(t_2) \subset \mathbb{R}^{n+1}$  varies continuously with  $t_2$  if given an  $\epsilon > 0$  there exists a  $\delta > 0$  so that for  $|t_2 - t_1| < \delta$ ,

$$d(\hat{K}(t_1), \hat{K}(t_2)) < \epsilon$$

Lemma 1 Consider the system  $\hat{f}$  as above with attainable set  $\hat{K}(t_1) \subset \mathbb{R}^{n+1}$ . Then  $\hat{K}(t_1)$  varies continuously with  $t_1 < \infty$ .

Proof We need only show that each point  $\hat{x}(t_1)$  of  $\hat{K}(t_1)$  is close to some point  $\hat{x}(t_2)$  of  $\hat{K}(t_2)$  and conversely. That is, we need show that given  $\epsilon > 0$  there exists a  $\delta > 0$  so that when  $|t_1 - t_2| < \delta$  there exists  $\hat{x}(t_1) \in \hat{K}(t_1)$  such that  $|x(t_1) - x(t_2)| < \epsilon$  for each  $\hat{x}(t_2) \in \hat{K}(t_2)$  and conversely.

Let  $u_1(t)$  be an admissible controller on  $[t_0, t_1+1]$  and  $\hat{x}_1(t)$  the corresponding response. For  $t_1 \leq t_2 \leq t_1 + 1$  calculate

$$x_1^o(t_2) - x_1^o(t_1) = \int_{t_0}^{t_2} F(x_1(t))dt - \int_{t_0}^{t_1} F(x_1(t))dt$$

and

$$\begin{aligned} x_1(t_2) - x_1(t_1) &= \Phi(t_2) \int_{t_0}^{t_2} \Phi(s)^{-1} B(s)u_1(s)ds \\ &- \Phi(t_2) \int_{t_0}^{t_1} \Phi(s)^{-1} [B(s)u_1(s)]ds \\ &+ [\Phi(t_2) - \Phi(t_1)] \left[ \int_{t_0}^{t_1} \Phi(s)^{-1} B(s)u_1(s)ds \right]. \end{aligned}$$

So

$$x_1^o(t_2) - x_1^o(t_1) = \int_{t_1}^{t_2} F(x_1(t))dt$$

and

$$x_1(t_2) - x_1(t_1) = \Phi(t_2) \int_{t_1}^{t_2} \Phi(s)^{-1} u_1(s) ds \\ + [\Phi(t_2) - \Phi(t_1)] \left[ \int_{t_0}^{t_1} \Phi(s)^{-1} B(s) u_1(s) ds \right]$$

Since  $A(t)$  is bounded and continuous on  $[t_0, t_1+1]$  so is  $\Phi(t)$  and therefore there exists a constant  $C_1$  so that

$$|\Phi(t)| < C_1$$

and

$$|\Phi(t)^{-1}| < C_1 \text{ on } [t_0, t_1+1].$$

Also since  $B(s)$  has bounded continuous elements  $b_j^1(t)$  and  $u_1(t)$  is bounded and measurable there exists the constant  $C_2$  so that

$\left| \int_{t_0}^{t_1} \Phi(s)^{-1} B(s) u_1(s) ds \right| < C_2$ . Integration is a continuous operation, therefore, given an  $\epsilon > 0$  there exists a  $\delta > 0$  so that

$$\left| \int_{t_1}^t F(x_1(t)) dt \right| < \frac{\epsilon}{3},$$

$$\left| \int_{t_1}^t \Phi(s)^{-1} B(s) u_1(s) ds \right| < \frac{\epsilon}{3C_2}$$

for  $|t - t_1| < \delta < 1$ .

Hence

$$|\hat{x}_1(t_2) - \hat{x}_1(t_1)| < \frac{\epsilon}{3} + c_1 \frac{\epsilon}{3c_1} + \frac{\epsilon}{3c_2} c_2 = \epsilon$$

for  $|t_2 - t_1| < \delta < 1$ .

The other way we consider  $u_1(t) = u(t)$  on  $[t_0, t_1]$  where  $u(t)$  steers to  $\hat{x}(t_1)$  and extend it to  $[t_0, t_1+1]$  by letting  $u_1(t) = u(t_1)$  for  $t \in [t_1, t_1+1]$ . The above calculation is then repeated to find  $|\hat{x}(t_2) - \hat{x}(t_1)| < \epsilon$  for  $|t_2 - t_1| < \delta < 1$  and so  $\hat{K}(t_1)$  varies continuously with  $t_1$ .

Theorem 3 Consider the system  $\hat{f}$  as above with initial data  $\hat{x}_0 = (0, x_0)$ , compact restraint set  $\Omega$ , and set of attainability  $\hat{K}(t_1)$ . Let the target set  $G = \{x^0, x \mid 0 \leq x^0 \leq \beta, x \in \tilde{G}\}$  where  $\beta > 0$  is a constant and  $\tilde{G}$  is a compact set of  $R^n$ . Suppose  $G$  meets the interior of  $\hat{K}(t_1)$ , then there is a  $\delta > 0$  such that  $G$  meets  $\hat{K}(t_1)$  for  $|t - t_1| < \delta$ .

Proof Since  $G$  meets the interior of  $\hat{K}(t_1)$ , there is a point  $\hat{p} \in (G \cap \text{Int. } \hat{K}(t_1))$  and a ball neighborhood  $N(\hat{p})$  of radius  $r > 0$  contained in  $\hat{K}(t_1)$ . Consider the hyperplane  $x^0 = p^0 - r/2$  of  $R^{n+1}$  and in this plane pick  $n+1$  independent points  $\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n, \hat{x}_{n+1}$  of the boundary of the ball  $N(\hat{p})$ , all equally spaced. Let  $\hat{x}_1(t), \hat{x}_2(t), \dots, \hat{x}_n(t), \hat{x}_{n+1}(t)$  be responses of  $\hat{f}$  with initial data  $\hat{x}_0 = (0, x_0)$  and corresponding to controllers  $u_1(t), u_2(t), \dots, u_{n+1}(t)$ ,  $t_0 \leq t \leq t_1 + 1$ , which are such that  $\hat{x}_1(t_1) = \hat{x}_1, \dots, \hat{x}_{n+1}(t_1) = \hat{x}_{n+1}$ . Pick  $1 > \delta > 0$  so small that for  $|t - t_1| \leq \delta$  the points  $\hat{x}_1(t)$  lie within spheres of radius  $r/10$  of the points  $\hat{x}_1, \dots, \hat{x}_{n+1}$ . This being possible because of the previous lemma 1.

Consider the convex combination of controllers  $u_\lambda(t) = \lambda_1 u_1(t) + \lambda_2 u_2(t) + \dots + \lambda_{n+1} u_{n+1}(t)$ ,  $\lambda_i \geq 0$ ,  $\sum \lambda_i = 1$  (Note  $|u_\lambda| \leq 1$ ) and the corresponding responses  $\hat{x}_\lambda(t)$  of  $\hat{f}$  with initial data  $(0, x_0)$ . For each fixed  $t$ ,  $|t - t_1| \leq \delta$  these response end points  $x_\lambda(t)$  sweep out a surface section  $\tilde{S}$  which lies below the plane  $x^0 = p^0$  by convexity, above or on the plane  $x^0 = 0$  because of the positive nature of  $F$  and intersect the line segment  $\{0 \leq x^0 \leq p^0, x = p\}$  (see proof of theorem 1). Hence  $G$  meets  $\hat{K}(t)$  for  $|t - t_1| \leq \delta < 1$ .

We now consider the problem of existence of optimum controllers.

Theorem 4 Consider the system  $\hat{f}$  as above with compact restraint set  $\Omega = \{u \mid |u^i| \leq 1, i=1, 2, \dots, m\} \subset R^m$ , initial point  $(0, x_0) \in R^{n+1}$  at time  $t_0$  and constant compact target set  $G = \{x^0, x \mid 0 \leq x^0 \leq \beta, x \in \tilde{G}\}$  for  $\beta > 0$ . If there exists an admissible controller  $u(t) \in \Omega$  steering  $\hat{x}_0$  to  $G$  on  $t_0 \leq t \leq t_1$  then there exists an optimum controller (also admissible) steering  $\hat{x}$  to  $G$  in minimum time duration  $t^* - t_0$ .

Proof If  $(0, x_0) \in G$  then  $t^* = t_0$  and optimum control is not required. So assume  $(0, x_0) \notin G$  and consider the set of attainability  $\hat{K}(t_1)$  for  $t_1 \geq t_0$ . Since there is one controller which steers  $(0, x_0)$  to  $G$  the set  $\hat{K}(t_1)$  meets  $G$  for some  $t_1 > t_0$ . Define  $t^*$  to be the greatest lower bound of all times  $t_1$  such that  $\hat{K}(t_1)$  meets  $G$ . By the continuous dependence of  $\hat{K}(t_1)$  on  $t_1$  the set of times for which  $K(t_1)$  meets  $G$  is a closed set in  $R^1$ . Hence  $t^*$  is the first time  $\hat{K}(t_1)$  meets  $G$  and therefore pick as the optimum controller  $u^*(t)$ ,  $t_0 \leq t \leq t^*$ , a controller which steers to

$$K(t^*) \cap G.$$

The next theorem asserts that for optimum control we need only consider points of the lower boundary of the set of attainability and therefore by theorem 2 extremal controllers.

A sufficiency condition is also included.

Theorem 5. Consider the system  $\hat{f}$  as above with compact rectangular restraint set  $\Omega$ , initial point  $(0, x_0)$  at  $t_0$  and compact convex target set  $G = \{x^0, x | 0 \leq x^0 \leq \beta; x \in G; \beta > 0\}$ . Let  $u^*(t)$  be a minimal time optimal controller steering  $\hat{x}^*(t)$  from  $\hat{x}_0$  to  $G$ . Then  $u^*(t)$  is extremal, that is, there exists a nonvanishing adjoint response  $\hat{\eta}(t) = (\eta_0, \eta(t))$  with  $\eta_0 \leq 0$  so that

$$\eta(t)B(t)u^*(t) = \text{Max}_{u \in \Omega} \{ \eta(t)B(t)u \}$$

almost always on  $[t_0, t^*]$  with  $\hat{\eta}(t^*)$  an outward normal of  $\hat{K}(t^*)$  at  $\hat{x}^*(t^*)$  on  $\partial\hat{K}(t^*)$  and  $\hat{\eta}(t^*)$  satisfies the transversality condition, namely,  $\hat{\eta}(t^*)$  is normal to a supporting hyperplane  $\pi$  of  $G$  and the set of attainability  $\hat{K}(t^*)$  which separates  $\hat{K}(t^*)$  from  $G$ .

Moreover, if for each point [3]  $\bar{x} \in G$  there exists a nonmaximal controller  $\bar{u}(t) \in \Omega$  so that on  $\bar{t}_0 \leq t < \infty$  the response  $x_{\bar{u}}(t)$  initiating at  $\bar{x} = x_{\bar{u}}(\bar{t}_0)$  is contained in  $G$ , then when  $u(t)$  is an admissible extremal controller steering  $x_0$  to  $G$  by means of a response satisfying the transversality condition it is an optimum controller.

Proof By assumption there exists a controller steering  $\hat{x}_0$  to  $G$  so  $G$  meets  $\hat{K}(t^*)$ . Suppose  $G$  meets the interior of  $\hat{K}(t^*)$ . This is impossible because then  $G$  meets the interior of  $\hat{K}(t)$  for  $|t - t^*| < \delta$ ,  $\delta > 0$ , by theorem 3 and this contradicts the optimality of the controller. Hence  $\partial G$  meets  $\partial\hat{K}(t^*)$  so that the optimum controller must steer to  $\partial\hat{K}(t^*)$ . We must show that it steers to a lower boundary point to conclude that it is extremal. This follows at once because  $\hat{K}(t)$  always first makes contact with  $G$  at a lower boundary



point as can be seen by considering how the compact set  $\hat{K}(t_1)$  with convex lower surface moves with respect to the set  $G$ . Thus if  $u^*(t)$  is optimal it is extremal and by theorem 2 there exists the nonvanishing adjoint response  $\hat{\eta}(t)$  so that

$$\eta(t)B(t)u^*(t) = \text{Max}_{u \in \Omega} \eta(t)B(t)u$$

where  $\hat{\eta}(t^*)$  satisfies the transversality condition since  $G$  and the lower boundary of  $\hat{K}(t^*)$  are convex they can be separated by a supporting hyperplane  $\pi$  and we choose  $\hat{\eta}(t^*)$  to be normal to  $\pi$  and directed into the halfspace containing  $G$ .

When  $u(t)$  is an admissible extremal controller steering  $\hat{x}_0$  to  $G$  and satisfying the transversality condition it must be an optimum controller if  $G$  has the property that through each point  $\bar{x} \in G$  there passes a nonmaximal response which remains forever in  $G$ . This follows because once  $G$  and  $\hat{K}(t)$  come together the interior of  $\hat{K}(t)$  has a nonempty intersection with  $G$  so that the transversality condition can only be satisfied once and therefore there is only one time, namely  $t^*$ , for which an extremal controller can steer to  $G$  and satisfy the transversality condition. Thus any such extremal controller satisfying the transversality condition is an optimum controller.

Q.E.D.

We have therefore reduced the problem of finding an optimum controller for the approximation problem to that of finding a solution to the two point boundary value problem as given by the  $2n+2$  equations:

$$\dot{x}^0 = F(x)$$

$$\dot{x} = A(t)x + B(t) \max_{u \in \Omega} \{ \eta(t)B(t)u \}$$

$$\dot{\eta} = -\eta A(t) - \eta_0 \frac{\partial F'}{\partial x}(x)$$

$$\dot{\eta}_0 = 0 \quad (\eta_0 \leq 0)$$

with boundary conditions  $\hat{x}(t_0) = \hat{x}_0$ ,  $\hat{x}(t^*) \in \partial G$  with  $\hat{\eta}(t^*)$  an interior normal to  $G$  at  $\hat{x}(t^*)$ .

### 3) An Example of Approximate Bounded Phase Coordinate Time Optimal Control

We shall consider a very simple example to illustrate some of the theory of the previous section. Consider a simple mechanism with position coordinate  $x$  and velocity coordinate  $y$ . Suppose it is desired to bring the mechanism to rest by means of a thrust force  $u(t)$  whose magnitude is bidirectional but limited to be less than 1 in magnitude and suppose the velocity is not to exceed .6 in magnitude. That is, consider the linear system

$$\dot{x} = y$$

$$\dot{y} = u(t)$$

with  $|u(t)| \leq 1$ ,  $\Lambda = \{x, y \mid |y| \leq .6\}$ ,  $x(0) = 10$ , and  $y(0) = 0$ .

$$\begin{aligned}
\text{Pick } F(x,y) &= \frac{1}{2}(y - \frac{1}{2})^2 & \text{for } y \geq \frac{1}{2} \\
&= 0 & \text{for } |y| \leq \frac{1}{2} \\
&= +\frac{1}{2}(y + \frac{1}{2})^2 & \text{for } y \leq -\frac{1}{2}
\end{aligned}$$

We shall later determine the parameter  $\beta > 0$  so that the strict bound on  $y$  is not exceeded. Problems in which the bound is soft are more easily handled since then we can generally pick  $\beta$  ahead of time and in a straightforward manner solve the two point boundary value problem. Here we have picked  $F(x,y)$  so that we are constraining the response even before the boundary of  $\Lambda$  is exceeded in hopes of maintaining the strict bound on  $y$ . To solve this approximate problem it is merely required that we find a solution of the system:

$$\dot{x}^0 = F(x,y)$$

$$\dot{x} = y$$

$$\dot{y} = \text{Max}_{u \in \Omega} \{\eta_2 u\}$$

$$\dot{\eta}_0 = 0 \quad (\eta_0 \leq 0)$$

$$\dot{\eta}_1 = 0$$

$$\dot{\eta}_2 = -\eta_1 - \eta_0 \frac{\partial F}{\partial y}$$

with  $x^0(0) = 0$ ,  $x(0) = 10$ ,  $y(0) = 0$ ,  $x^0(t_1) \leq \beta$ ,  $x(t_1) = 0$ ,  $y(t_1) = 0$  for some  $t_1 > 0$ .

A simple calculation shows that picking  $\delta = .08$ ,  $\eta_0(0) = -10$ ,  $\eta_1(0) = -1$ ,  $\eta_2(0) \approx -.55$  provides a time optimal solution for this problem. A plot of this response is given by figure 1. Note in this problem the exact optimum solution was obtained, but in general one would pick different  $F(x,y)$ 's to get better approximations.

#### 4) Remarks on the approximate bounded phase coordinate problems with integral cost

As before consider the linear control process

$$\dot{x} = A(t)x + B(t)u(t)$$

satisfying the conditions stated at the beginning of section 1. As a cost functional of control consider

$$C(u) = g(x(T)) + \int_{t_0}^T \{f^0(x,t) + h^0(u,t)\}dt$$

where  $T = \text{fixed time} > t_0$  and the real functions  $f^0(x,t)$  and  $h^0(u,t)$  are continuously differentiable and  $f^0(x,t)$  is a convex function of  $x$  for each  $t$ .

The problem of optimal control is to pick an admissible controller  $u(t)$  on  $[t_0, T]$  so that the response  $x_u(t)$  of  $\dot{x}$  moves from  $x_0$  to a target set  $\tilde{G} \subset \mathbb{R}^n$  at  $T$ , ( $\tilde{G}$  may be whole space) and minimizes  $C(u)$  with the entire response  $x_u(t)$  contained in the closed convex restraint set  $\Lambda$ .

As before we introduce the convex differentiable function  $F(x)$  satisfying the conditions

$$\begin{aligned} F(x) &> 0 \quad \text{if } x \notin \Lambda \\ &= 0 \quad \text{if } x \in \Lambda \end{aligned}$$

The approximation problem is obtained by adding  $F(x)$  to the integrand of the cost functional  $C(u)$  to obtain a new cost functional

$$\begin{aligned} C_\lambda(u) &= g(x(T)) + \int_{t_0}^T \{f^0(x,t) + \lambda F(x) + h^0(u,t)\}dt \\ &= \int_{t_0}^T \{\tilde{f}^0(x,t) + h^0(u,t)\}dt, \end{aligned}$$

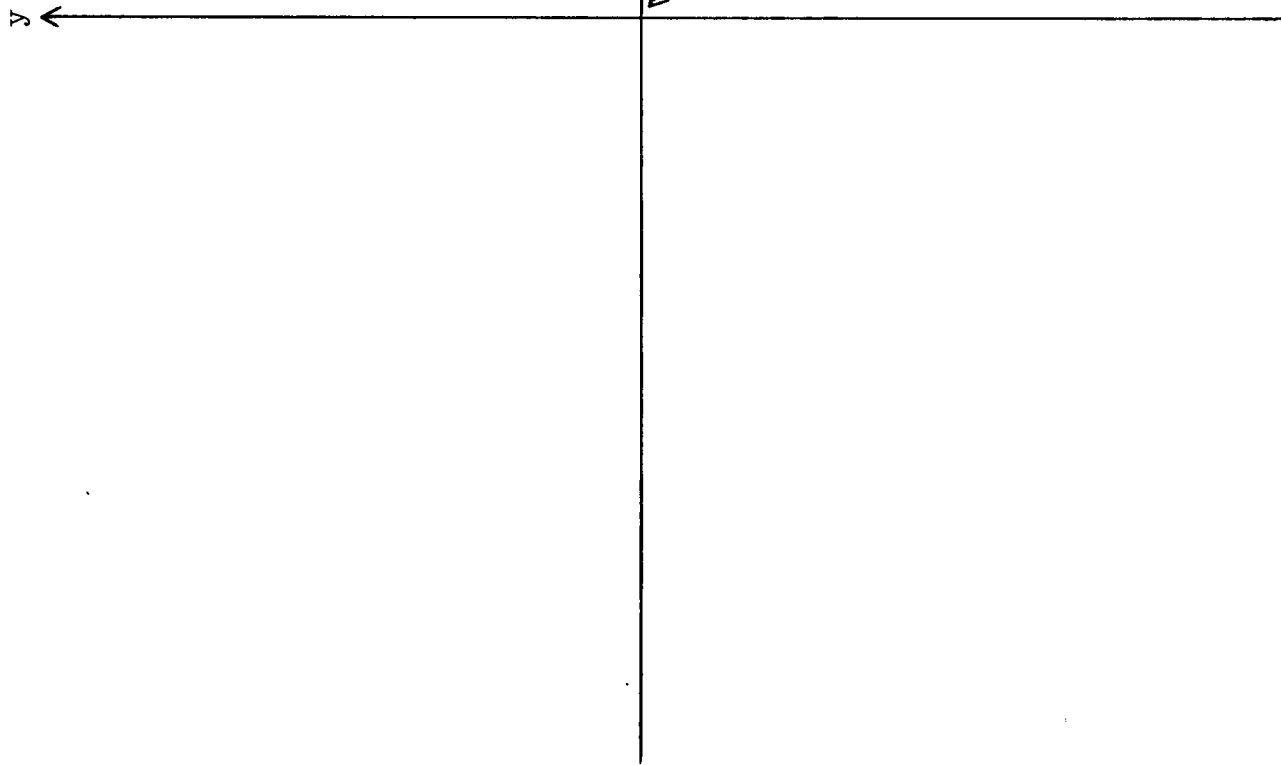
here  $\lambda \geq 0$ . If  $\lambda$  is sufficiently large then one would expect that the contribution from the term  $\lambda F(x)$  can be small only if the response stays near  $\Lambda$  or within it. The approximation problem is to find that controller  $u(t)$  which minimizes  $C_\lambda(u)$  and steers to  $\tilde{G} \subset R^n$ .

We shall assume that  $h^\circ(u, t)$  is convex in  $u$  for each  $t$  or that the controller is bounded and  $h$  is a positive function of  $u$  for each  $t$ . In either case the previous theory can be applied after slight modification by noting that  $\tilde{f}^\circ(x, t) = f^\circ(x, t) + \lambda F(x)$  is a convex function of  $x$  for each  $t$  since both  $f^\circ$  and  $F$  were convex functions and by noting the contribution to  $x^\circ(T)$  made by the terms  $h^\circ(u, t)$ . That is, the problem has now been cast as one which is covered by the sufficiency results of reference 5 which are also necessary [reference 7] and can be obtained as a slight modification of the results of section 2.

## References

1. Chang, S. S. L., "Optimal control in bounded phase space," *Automatica*, Vol. 1, (1962), 55-57.
2. Gamkrelidze, R. V., "Optimal processes with bounded phase coordinates," *Izv. Akad. Nauk. SSSR, Ser. Mat.*, Vol. 24, (1960), 315-356.
3. Harvey, C. A., and Lee, E. B., "On the Uniqueness of Time-Optimal Control," *J. Math. Anal. and App.*, Vol. 5, (1962), 258-268.
4. LaSalle, J. P., "The time optimal control problem," *Am. Math. Studies No. 45*, (1960), 1-24.
5. Lee, E. B., "A sufficient condition in the theory of optimal control," *SIAM Journal of Control*, Vol. 1., No. 3, (1963), 241-245.
6. Lee, E. B., and Markus, L., "Optimal Control for Nonlinear Processes," *Archive for Rational Mechanics and Analysis*, Vol. 8, No. 1, (1961), 36-58.
7. Pontryagin, L. S., Boltayanskii, V. G., Gamkrelidze, R. V., Mishehenko, E. F., "The Mathematical Theory of Optimal Processes," John Wiley and Sons, New York, 1962.
8. Roxin, E., "The existence of optimal controls," *Mich. Math. Journal*, Vol. 9, (1962), 109-119.
9. Russell, D., "Time Optimal bounded phase coordinate control of linear systems parts I, II, III, Appendix C of Honeywell MPG Report 12006-QR 1 (The First Quarterly Progress Report on Control NASw-986)," 30 September, 1964.

Institute of Technology  
University of Minnesota  
Minneapolis, Minnesota



$$\dot{x}^0 = F(x, y)$$

$$\dot{x} = y$$

$$\dot{y} = u(t) \quad |u(t)| \leq 1$$

$$\begin{aligned} F(x, y) &= \frac{1}{2}(y - \frac{1}{2})^2 & \text{for } y \geq \frac{1}{2} \\ &= 0 & \text{for } |y| \leq \frac{1}{2} \\ &= \frac{1}{2}(y + \frac{1}{2})^2 & \text{for } y \leq -\frac{1}{2} \end{aligned}$$

NOTES ON THE RESTRICTED THREE BODY PROBLEM:  
APPROXIMATE BEHAVIOR OF SOLUTIONS NEAR THE COLLINEAR  
LAGRANGIAN POINTS

C. C. Conley

Introduction

The purpose of these remarks is to describe in some detail the geometry of solutions of the restricted three body problem (as viewed in the rotating coordinate system) near those equilibrium points which are collinear with the two positive masses.

We deal only with the linearized equations, but make some qualitative observations which can be carried over without difficulty to the nonlinear equations for suitable values of the Jacobi Constant.

This report is intended to be the first in a series whose ultimate aims include an existence proof for the "periodic" solutions discovered numerically by M. Davidson [1]. Whether or not this can be accomplished remains to be seen, but it does seem clear that a thorough understanding of the behavior of orbits near the equilibrium point will be required. More will be said about this question in later reports.

From the work in this report we obtain the following qualitative picture of solutions of the linearized equations for values of the "Jacobi Constant" slightly above that of the equilibrium point.

The projections of orbits into the configuration space are constrained to lie in the region  $R$  between the two branches of a hyperbola symmetric with respect to the line,  $\ell$ , joining the positive mass points, which line is contained in  $R$ .

We will generally restrict our attention to the portion of the phase space corresponding to a closed interval  $I$  of  $\ell$  about the projection of the equilibrium point. Recalling that the value of the integral is fixed, we will see that this portion of the phase space is homeomorphic to  $S^2 \times I$  ( $S^2$  is the two-sphere) and so may be viewed as the space between 2-concentric spheres together with the bounding spheres.

N65 33058



If  $I$  is large enough we will see there there is exactly one closed orbit in this portion of the phase space. This corresponds to one of the family of periodic solutions which are known (by a theorem of Lyapounov) to exist in a neighborhood of the equilibrium point even for the nonlinear equations.

There are four "cylinders" in the phase space which abut on this periodic orbit and which are invariant under the flow. Two of these run to the outer bounding sphere and two to the inner. One of each of these two pair of cylinders corresponds to a family of solutions which is asymptotic to the periodic solution as the time goes to  $+\infty$ ; the others to families asymptotic as time goes to  $-\infty$ . These cylinders act as separatrices. They separate those solutions which go from the inner to the outer sphere (or vice versa) from those that do not: in the language of the configuration space, they separate those solutions which make a transit of the region of the equilibrium from those which do not cross this region. (The existence of such cylinders for the restricted problem is apparent. From a theorem of J. Moser [2] it can be seen that they are described by real analytic functions near the equilibrium point.)

The projection of these cylinders into the configuration space covers the union of two infinite strips the boundaries of which are the enveloping lines of the solutions asymptotic to the periodic solution (figure 1). These four enveloping lines (which are tangent to the hyperbolas bounding  $R$  as well as to the periodic orbit) divide  $R$  into several regions and we will be able to determine the nature of solutions in these different regions. Further description will be easier to give later.

An amusing result is that exactly one solution from each of the four cylinders of solutions asymptotic to the periodic solution has a cusp (as viewed in the configuration space). A modification of this statement holds as well for the restricted three body problem. These four cusp points determine arcs on the hyperbolas bounding  $R$ , and any solution which cusps on these arcs is making a transit of the equilibrium region.

A statement which is perhaps a little more useful is that there are two unique solutions which are "best" for making a transit of the equilibrium region in that they take the least time. One of the (possible) difficulties in using orbits which correspond to the solutions of M. Davidson is the amount of time it is possible to spend in the region of the equilibrium. \* It may be useful to have a simple criterion for decreasing

---

\* The values of the Jacobi Constant considered here are small relative to the ones usually considered.

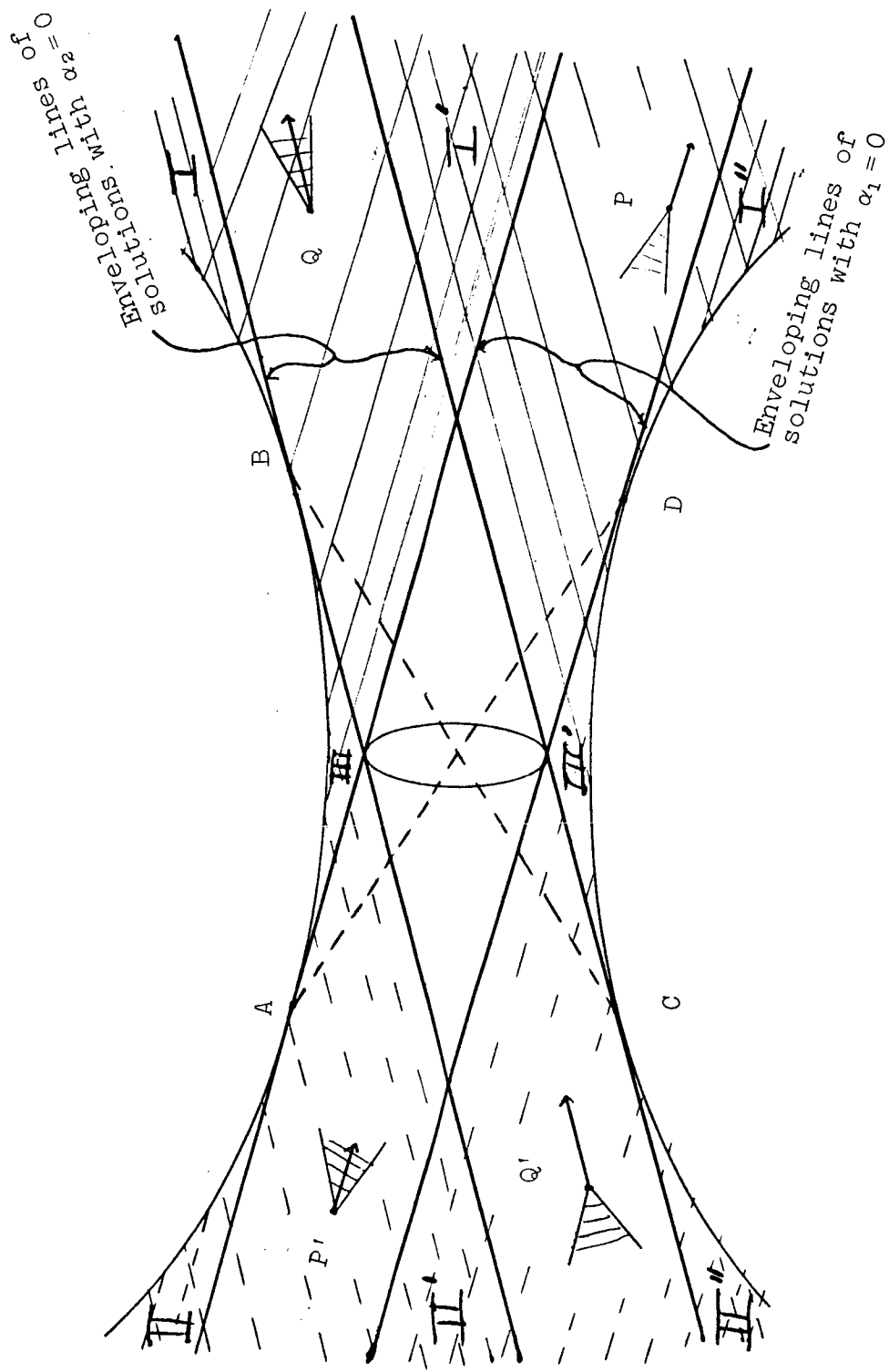


FIGURE 1. SOLUTIONS WITH VELOCITY VECTOR IN SHADED WEDGES GO ACROSS THE EQUILIBRIUM REGION

this time. An approximate means to determine the "best" orbit is given in statement eleven; a more accurate one could be derived using the result of J. Moser [2].

As stated above, these remarks have been collected primarily with a view to later applications. However, it is hoped they are of some value in themselves in gaining insight into the nature of solutions of the restricted three body problem.

## 1. The Equations

Without going through the arguments, we can state that the linearized equation near the equilibrium points in which we are presently interested form a hamiltonian system with Hamiltonian function:

$$(1) \quad H(x_1, x_2, y_1, y_2) = \frac{1}{2} \{ (y_1 - \omega x_2)^2 + (y_2 + \omega x_1)^2 - a x_1^2 + b x_2^2 \}$$

( $\omega, a, b$  are positive constants)

The equations are

$$(2) \quad \begin{aligned} \dot{x} &= Hy \\ \dot{y} &= -Hx. \end{aligned}$$

In these equations,  $\omega$  is the frequency of rotation of the coordinate system; we assume  $\omega$  is positive.

The constants  $a, b$  will be arbitrary positive constants in our discussion. In the case of the equilibrium point between the two positive masses of the restricted problem,  $a = 2b$ .<sup>\*</sup> If the mass ratio is that of the Earth and Moon, then with  $\omega = 1$ ,  $a$  is slightly larger than 8.

We introduce the following notation:

$$(3) \quad \begin{aligned} \hat{u} &= (x_1, x_2, y_1, y_2) \\ S &= \begin{pmatrix} -a & 0 \\ 0 & b \end{pmatrix}; \quad J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}; \quad I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}; \end{aligned}$$

---

<sup>\*</sup> This statement is also true of the other two equilibria considered, however, the next is not.

$$\mathcal{J} = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \quad (3)(\text{cont.})$$

$$\Sigma = \begin{pmatrix} \omega^2 I + S & \omega J \\ -\omega J & I \end{pmatrix}$$

Our equations are then written as

$$\begin{aligned} \hat{H}(\hat{u}) &= \frac{1}{2} (\hat{u}, \Sigma \hat{u}) \\ \dot{\hat{u}} &= \mathcal{J} \hat{H}_{\hat{u}} = \mathcal{J} \Sigma \hat{u}. \end{aligned} \quad (4)$$

Now to make the computations easier we introduce the non-canonical transformation

$$\begin{aligned} \hat{u} &= Au \\ A &= \begin{pmatrix} I & 0 \\ \omega J & I \end{pmatrix} \end{aligned} \quad (5)$$

The equations then transform to:

$$\begin{aligned} \dot{u} &= Bu \\ B &= A^{-1} \mathcal{J} \Sigma A = \begin{pmatrix} 0 & I \\ -S & -2\omega J \end{pmatrix} \end{aligned} \quad (6)$$

and the integral is given by

$$\begin{aligned} H(u) &= \hat{H}(A\hat{u}) = \frac{1}{2} (u, Eu) \\ E &= A^T \Sigma A = \begin{pmatrix} S & 0 \\ 0 & I \end{pmatrix} \end{aligned} \quad (7)$$

If we now write  $u = (x_1, x_2, z_1, z_2)$ , the equations above give  $\dot{x}_1 = z_2$ . Thus if we consider projections of orbits in the x-plane,  $z = (z_1, z_2)$  corresponds to the tangent vector.

In this notation we have for the integral:

$$H(u) = \frac{1}{2} (z_1^2 + z_2^2 - a x_1^2 + b x_2^2)$$

## 2. The Phase Space

We will be primarily interested in those orbits for which

$$(8) \quad H = h > 0,$$

and will describe the projections of these orbits in the  $x$ -plane.

Statement 1. a) For  $H = h$ , the projected orbits are constrained to move in the region  $R$  given by

$$R: \quad -a x_1^2 + b x_2^2 \leq h.$$

b) If  $h \geq 0$   $R$  is a connected region, otherwise it has two components.

c) If  $h > 0$ , the phase space is homeomorphic to  $S^2 \times E^1$  ( $S^2$  is the 2-sphere,  $E^1$  the real line). We will be most interested in that part of the phase space for which  $|x_1| \leq c > 0$ . This region can be considered as the space between two concentric spheres including the boundaries.

Proof: Only part c) needs comment. To see this statement, consider the line  $x_1 = c_1$ . On this line we have

$$z_1^2 + z_2^2 + b x_2^2 = 2h + a c_1^2$$

So the corresponding points in the phase space form a 2-sphere. The rest follows.

## 3. Computations

Statement 2. a) The matrix  $B$  has one pair of real eigenvalues and one pair of imaginary eigenvalues. These we denote by

$$\pm \mu, \pm i\nu \text{ where } \mu, \nu > 0.$$

b) The corresponding eigenvectors can be chosen to be:

$$\begin{array}{cccc} \mu & -\mu & i\nu & -i\nu \\ \hline v_1 = \begin{pmatrix} 1 \\ \sigma \\ \mu \\ \mu\sigma \end{pmatrix} & v_2 = \begin{pmatrix} 1 \\ -\sigma \\ -\mu \\ \mu\sigma \end{pmatrix} & w_1 = \begin{pmatrix} 1 \\ i\tau \\ i\nu \\ -\nu\tau \end{pmatrix} & w_2 = \bar{w}_1 = \begin{pmatrix} 1 \\ -i\tau \\ -i\nu \\ -\nu\tau \end{pmatrix} \end{array}$$

where  $\sigma$  and  $\tau$  are real,  $\sigma > 0$ ;  $\tau < 0$  (cf. e) of this Statement)

c) The general solution is of the form

$$u(t) = \alpha_1 e^{\mu t} v_1 + \alpha_2 e^{-\mu t} v_2 + 2 \operatorname{Re} (\beta e^{i\nu t} w_1)$$

where  $\alpha_1, \alpha_2$  are real,  $\beta$  is complex.

d) The value of the integral on the solution is

$$\frac{1}{2} (u(t), E u(t)) = \alpha_1 \alpha_2 e_1 + |\beta|^2 e_2$$

where

$$e_1 = (v_1, E v_2)$$

$$e_2 = (w_1, E w_2)$$

(Note: the inner product is the real one even when vectors are complex.)

e) The constants  $\mu, \nu, \sigma, \tau, e_1, e_2$  satisfy:

$$1) a - 2\omega\sigma\mu = \mu^2; \text{ in particular, } \mu > 0$$

$$2) -b\sigma + 2\omega\mu = \mu^2\sigma$$

$$3) a + 2\omega\tau\nu = -\nu^2; \text{ in particular, } \tau < 0$$

$$4) -b\tau + 2\omega\nu = -\nu^2\tau$$

$$5) (v_1, E v_1) = -a + b\sigma^2 + \mu^2 + \sigma^2\mu^2 = 0$$

$$6) (w_1, E w_1) = -a - b\tau^2 - \nu^2 + \nu^2\tau^2 = 0$$

$$7) (v_1, E w_1) = -a + ib\tau\sigma + i\mu\nu - \sigma\mu\tau\nu = 0$$

$$8) e_1 \equiv (v_1, E v_2) = -a - b\sigma^2 - \mu^2 + \sigma^2\mu^2 \\ = -2(b\sigma^2 + \mu^2) < 0$$

$$9) e_2 \equiv (w_1, E w_2) = -a + b\tau^2 + \nu^2 + \nu^2\tau^2 \\ = 2(b\tau^2 + \nu^2) > 0$$

$$10) 2ab(\sigma^2 + \tau^2) = e_2(a + b\sigma^2)$$

$$11) \sigma\mu\tau\nu = a ; \quad b\tau\sigma = -\mu\nu \quad (\text{from 7))}$$

$$\mu^2\nu^2 = ab; \quad -\tau\sigma = \sqrt{\frac{a}{b}}$$

Proof of Statement 2: (Recall  $B = \begin{pmatrix} 0 & I \\ -S & -2\omega J \end{pmatrix}$ )

To prove parts a) and b) and equations 1) - 4) of e), we first observe that any eigenvector must have a non-zero first component which we can take to be 1. The form of  $B$  then forces the eigenvector to be  $u = \{1, \rho, \lambda, \rho\lambda\}$  where  $\lambda$  is the eigenvalue. Now the last two equations in the system  $Au = \lambda u$  require that

$$a - 2\omega\lambda\rho = \lambda^2$$

$$-b\rho + 2\omega\lambda = \lambda^2\rho$$

Elimination of  $\rho$  gives

$$\lambda^4 + (b-a + 4\omega^2)\lambda^2 - ab = 0$$

and parts a) and b) as well as the first two equations of part e) follow.

Part c) needs no comment.

Parts d) and e) follow from general considerations:

Lemma 1. Let  $v$  and  $w$  be eigenvectors of the matrix  $\int \Sigma$  where  $\Sigma$  is symmetric and  $\int$  is skew symmetric and orthogonal, and let the corresponding eigenvalues be  $\lambda$  and  $\mu$  respectively.

Then either

$$(v, \Sigma w) = 0,$$

or

$$\lambda + \mu = 0$$

Proof: Since  $\int$  is orthogonal,

$$(v, \Sigma w) = (\int v, \int \Sigma w) = \mu(\int v, w)$$

$$(\Sigma v, w) = (\int \Sigma v, \int w) = \lambda(v, \int w).$$

The result follows by the symmetry of  $\Sigma$  and skew symmetry of  $J$ .

To apply this lemma to our problem we use the fact that, since  $B = A^{-1} J \Sigma A$ , the vectors  $Av_i$  and  $Aw_i$  are eigenvectors of  $J \Sigma$ . (The notation is that of §1.)

Part d) and equations 5) through 9) of e) now follow. The remaining equations and statements in e) are proved with a little algebra. The harder ones will be seen geometrically later so the computations are omitted.

Statement 3. If  $u(t)$  is a solution such that  $u(0) = (x_1, x_2, z_1, z_2)$ , then the constants  $\alpha, \beta$  (Statement 2, c) are given by:

$$e_1 \alpha_1 = -ax_1 - bx_2 - \mu z_1 + \mu \sigma z_2 = (u, E v_2)$$

$$e_1 \alpha_2 = -ax_1 + bx_2 + \mu z_1 + \mu \sigma z_2 = (u, E v_1)$$

$$e_2 \beta = -ax_1 - ib\tau x_2 - iv z_1 - v\tau z_2 = (u, E w_2)$$

Proof: This follows on dotting the equation

$$u(0) = \alpha_1 v_1 + \alpha_2 v_2 + \beta w_1 + \bar{\beta} w_2$$

with

$$E v_2, E v_1, E w_2 \text{ respectively, and using 5) - 9) of Statement 1).}$$

Statement 4. (Recall that

$$\begin{aligned} H(u) &= \frac{1}{2} \{z_1^2 + z_2^2 - ax_1^2 + bx_2^2\} \\ &= \alpha_1 \alpha_2 e_1 + |\beta|^2 e_2 \end{aligned}$$

where

$$e_1 < 0; e_2 > 0).$$

Consider the projection in the  $x$ -plane of solutions in the integral surface

$$H(u) = h > 0$$

The solutions in the integral surface divide into classes as follows:



1) The (unique) periodic solution:  $\alpha_1 = \alpha_2 = 0$

2) Solutions which are asymptotic to the periodic solution as  $t \rightarrow \infty$  ( $t \rightarrow -\infty$ ):

$$\alpha_1 = 0 \quad (\alpha_2 = 0)$$

3) Solutions whose  $x_1$  component tends to  $+\infty$  ( $-\infty$ ) as  $t \rightarrow \pm \infty$ :

$$\alpha_1, \alpha_2 > 0 \quad (\alpha_1, \alpha_2 < 0).$$

These are solutions whose projected orbits in the  $x$ -space lie in a half space  $x_1 > c$  or  $x_1 < \tilde{c}$ . They do not make a "transit" of the equilibrium region.

4) Solutions whose  $x_1$  component goes from  $-\infty$  to  $+\infty$  ( $+\infty$  to  $-\infty$ ) as  $t$  goes from  $-\infty$  to  $+\infty$ :

$$\alpha_1 > 0, \alpha_2 < 0 \quad (\alpha_1 < 0; \alpha_2 > 0).$$

These are the solutions which do cross the equilibrium region.

Proof: By inspection of the corresponding general solution.

We are particularly interested in the solutions of class 4) which, in the case of the equilibrium between the two positive mass points, can be interpreted as solutions going from the earth side of the equilibrium to the moon side (or vice versa). Clearly the most "efficient" (least time expenditure) such orbit is that for which  $\beta = 0$  since the " $\beta$ -portion" of a solution contributes only useless oscillation — we will come back to this point later.

Interpretation for restricted Problem:

Solution 1) corresponds of course to the periodic solution about the equilibrium point of the restricted problem whose existence is guaranteed by a theorem of Lyapounov.

The solutions of 2) correspond to the four families which are asymptotic to the periodic solution as described in the introduction. Since the argument of  $\beta$  is free and can vary on a "circle," these four families

are easily seen to be "cylinders" which abut on the periodic solution. One can now check that two of these cylinders "go to  $+\infty$ " and two "to  $-\infty$ " (as  $t \rightarrow +\infty$ ), i. e., the region of the earth (say) or the moon (resp.) Again, one easily checks that one of each of these pairs is asymptotic to the periodic solution as  $t$  goes to  $+\infty$ , the other as  $t$  goes to  $-\infty$ .

The solutions of 3) are those which enter the region of the equilibrium only to return whence they came while those of 4) make the transit.

While we have considered only the linearized equations, simple considerations ensure the same qualitative picture for the equations of the restricted problem.

Statement 5. If  $x_1^2 > \frac{2h(a - \mu^2)}{a\mu^2} \equiv c^2$

then a)  $x_1 z_1 \geq 0 \Rightarrow \alpha_1 x_1 > 0$

b)  $x_1 z_1 \leq 0 \Rightarrow \alpha_2 x_1 > 0$

Interpretation: If a solution crosses the line  $x_1 = c_1$  going away from the origin, then if  $c_1 > c$ , the  $x_1$  component of this solution must tend to  $+\infty$ . If a solution crosses the line coming toward the origin, it's  $x_1$  component goes to  $+\infty$  as  $t \rightarrow -\infty$ . Corresponding statements hold if  $x_1 = c_1 < -c$ .

In particular, a solution of class 2) or 4) ( $\alpha_1 \alpha_2 \leq 0$ ) can cross the line  $x_1 = c$ , only once and must do so with  $z_1 \neq 0$ . We will make use of this remark later.

Also, we can see that a solution crosses both of the lines  $x_1 = \pm c_1$  if and only if  $\alpha_1 \alpha_2 < 0$ . This comment allows us to give a precise geometric meaning to the statement that "a solution makes a transit of the equilibrium region." A similar definition works for the restricted problem for the same reason.

Proof of Statement 5.

a) We have (Statement 3)

$$e_1 \alpha_1 = -ax_1 - bx_2 - \mu z_1 + \mu \sigma z_2,$$

where  $e_1 < 0$  (Statement 2), e), 9)

Thus

$$\operatorname{sgn} x_1 \alpha_1 = \operatorname{sgn}(ax_1^2 + b\sigma x_1 x_2 + \mu z_1 x_1 - \mu \sigma x_1 z_2)$$

Since  $x_1 z_1 \geq 0$ , we need only show that

$$ax_1^2 > |b\sigma x_1 x_2 - \mu \sigma x_1 z_2|$$

We estimate (Schwarz)

$$|b\sigma x_1 x_2 - \mu \sigma x_1 z_2| \leq |x_1| (b\sigma^2 + \mu^2 \sigma^2)^{\frac{1}{2}} (bx_2^2 + z_2^2)^{\frac{1}{2}}$$

Using Statement 2, e) 1) and the energy integral we have

$$a - \mu^2 = b\sigma^2 + \mu^2 \sigma^2$$

$$bx_2^2 + z_2^2 \leq 2h + ax_1^2$$

so that

$$\begin{aligned} |b\sigma x_1 x_2 - \mu \sigma x_1 z_2| &\leq |x_1| (a - \mu^2)^{\frac{1}{2}} (h + ax_1^2)^{\frac{1}{2}} \\ &= |x_1| (a^2 x_1^2 + 2ha - 2h\mu^2 - a\mu^2 x_1^2)^{\frac{1}{2}} \end{aligned}$$

This last quantity is less than  $ax_1^2$  provided  $2ha - 2h\mu^2 - a\mu^2 x_1^2 < 0$  which is the hypothesis. A similar proof holds for part b).

---

A statement stronger than the above can be proved if we place a restriction on the constants  $a$  and  $\omega$ : Namely

Statement 6. Recall the equations are given by

$$\dot{x}_1 = z_1 ; \quad \dot{z}_1 = -2\omega z_2 + ax_1$$

$$\dot{x}_2 = z_2 ; \quad \dot{z}_2 = 2\omega z_1 - bx_2,$$

If

$$x_1 = c_1 \sqrt{\left( \frac{8\omega^2 h}{a^2 - 4\omega^2 a} \right)}$$

Then  $z_1 \geq 0$  implies the corresponding solution never returns to the line  $x_1 = c_1$ . (Also  $x_1(t) \rightarrow \infty$ ) Furthermore if  $z_1 = 0$ ,  $c_1$  is an absolute minimum for  $x_1$ . A similar statement holds if

$$x_1 = c_1 < -\sqrt{\frac{8\omega^2 h}{a^2 - 4\omega^2 a}}$$

Proof: The proof consist of showing that  $\dot{z}_1 > 0$  under the above circumstances. We have:

$$|z_2| \leq \sqrt{2h + ax_1^2}$$

so that

$$4\omega^2 z_2^2 \leq 4\omega^2 (2h + ax_1^2) < a^2 x_1^2$$

The last inequality being the hypothesis. The result now follows.

This statement has no force unless

$$a^2 - 4\omega^2 a > 0$$

which situation does however hold for the equilibrium point of the restricted problem between the two positive masses. ( $a \geq 8$ ;  $\omega = 1$ ).

Geometrically, we see from Statement 6 that the points where the  $x_1$  component of a solution can have a maximum must lie to the left of the line  $x_1 = \sqrt{\frac{8\omega^2 h}{a^2 - 4\omega^2 a}}$ . Such a restriction is valid only when  $a^2 - 4\omega^2 a > 0$  as can easily be seen. This remark will be useful in a later report.

#### Statement 7

The projection of the periodic solution in the x-plane is an ellipse with minor axis of length  $2\sqrt{\frac{h}{e_2}}$  in the direction of the  $x_1$ -axis and major axis of length  $-2\tau\sqrt{\frac{h}{e_2}}$  in the direction of the  $x_2$ -axis.

Proof: (Assume  $\beta$  is real.) The projection is given by

$$x_1(t) = 2 \operatorname{Re}(\beta e^{i\nu t}) = 2\beta \cos \nu t$$

$$x_2(t) = -2\tau \operatorname{Im}(\beta e^{i\nu t}) = -2\tau \beta \sin \nu t$$

Also the energy integral gives

$$\beta^2 e_2 = h$$

That  $|\tau| > 1$  follows from Statement 2, e), 6):

$$0 < \tau^2 = \frac{\nu^2 + a}{\nu^2 - b} > 1.$$

The result follows.

Statement 8. (Recall the solutions with  $\alpha_1 \alpha_2 = 0$  are those asymptotic to the periodic solution.)

a) The envelopes of projections in x-space of orbits with  $\alpha_1 = 0$  are the straight lines

$$\begin{aligned} x_2 &= -\sigma x_1 \pm (a - \sigma^2 b) \sqrt{\frac{2h}{ab}} \\ &= -\sigma x_1 \pm 2 \sqrt{\frac{h}{e_2}} (\sigma^2 + \tau^2)^{\frac{1}{2}}. \end{aligned}$$

The corresponding envelopes for  $\alpha_2 = 0$  are:

$$x_2 = \sigma x_1 \pm (a - \sigma^2 b) \sqrt{\frac{2h}{ab}}.$$

b) All four of these lines are tangent to the boundaries of  $R$  (i. e., of the region of x-space wherein solutions must move — see Statement 1.)

c) The points of tangency lie on the lines

$$x_1 = \pm \sigma \sqrt{\frac{2bh(a - b\sigma^2)}{a}} = \mp \frac{1}{\tau} \sqrt{2h(a - b\sigma^2)}$$

(See figure 1)

#### Proof of Statement 8

a) If  $\alpha_1 = 0$  we have (Statement 2)

$$x_1 = \alpha_2 e^{-\mu t} + 2 \operatorname{Re}(\beta e^{i\nu t})$$

$$x_2 = -\sigma \alpha_2 e^{-\mu t} - 2\tau \operatorname{Im}(\beta e^{i\nu t})$$

$$= -\sigma x_1 + 2\sigma \operatorname{Re}(\beta e^{i\nu t}) - 2\tau \operatorname{Im}(\beta e^{-i\nu t})$$

The extreme values of  $x_2$  for fixed  $x_1$  are obtained by varying  $\arg \beta$ . These are computed to be

$$x_2 = -\sigma x_1 \pm 2|\beta|(\sigma^2 + \tau^2)^{\frac{1}{2}},$$

Finally, we have  $|\beta| = \sqrt{\frac{h}{e_2}}$  from the energy integral which gives one of the alternate expressions in a). Observe that the extreme values are achieved.

b) We could prove b) by computation; however, the following geometric argument carries over to the corresponding statement (that "envelopes of solutions asymptotic to the periodic solution touch the boundaries of  $R''$ ) for the restricted problem:

We first observe that we can obtain a space homeomorphic to the phase space as follows: First deform  $R$  to an infinite strip (i.e., squeeze the boundaries down to straight lines). Noting that at each point of  $R$  (except the boundaries) there is a "circle" of possible velocities (i.e.,  $z_1^2 + z_2^2 = \text{const} > 0$ ) we cross the infinite strip with a circle to obtain a "pipe," i.e., the space between two coaxial cylinders.

The length along the cylinder corresponds to the  $x_1$  coordinate. For each fixed  $x_1$  there corresponds an annulus of points; the radial variable in this annulus corresponds to  $x_2$ , while the angular variable corresponds to the direction of the velocity vector  $z = (z_1, z_2)$ . The inner and outer boundaries of the annulus correspond to boundary points of  $R$ . These boundaries should be identified to (different) points since  $z_1^2 + z_2^2$  is zero on the boundary of  $R$ ; however, we neglect this point for the moment.

Now consider the "cylinder" of solutions with  $\alpha_1 = 0$  say. For fixed  $x_1$ , the corresponding points on the cylinder make a closed loop in the pipe.

Now if  $x_1 > c$  (Statement 5), and  $\alpha_1 = 0$ , then  $z_1 < 0$ . Thus the

corresponding "circle" does not go around the hole in the pipe. On the other hand, the periodic orbit does encircle the hole since the velocity vector on this orbit goes through all angles. Since the cylinder abuts on this periodic orbit, some section of it must enclose the hole. It follows that this cylinder must cross one of the bounding cylinders of the pipe.

This implies that some orbit with  $\alpha_1 = 0$  must touch the boundary of  $R$  and so the envelopes of these orbits must cut this boundary. However, they cannot go out of the region  $R$ , and therefore are tangent to the boundary.

Part c) (and alternate expression in part a))

From parts a) and b) it follows that, for example, the equations

$$a x_1^2 - b x_2^2 + 2h = 0$$

$$x_2 = -\sigma x_1 + 2\sqrt{\frac{h}{e_2}} (\sigma^2 + \tau^2)^{\frac{1}{2}}$$

have a unique solution for  $x_1$ .

This means the following quadratic equation has double roots:

$$(a - b\sigma)x_1^2 + 4b\sigma\sqrt{\frac{h}{e_2}} (\sigma^2 + \tau^2)^{\frac{1}{2}} - 4b\frac{h}{e_2} (\sigma^2 + \tau^2) + 2h = 0.$$

The condition for a double root is:

$$4b^2\sigma^2\frac{h}{e_2} (\sigma^2 + \tau^2) = (a - b\sigma^2) \left\{ 2h - \frac{4bh}{e_2} (\sigma^2 + \tau^2) \right\}$$

which reduces to the equation:

$$(\sigma^2 + \tau^2) = \frac{e_2(a - b\sigma^2)}{2ab}$$

This equation (which is Statement 2, e), 10)) could of course be verified algebraically from the other equations of e); the algebra is left out since the geometric proof suffices.

The remaining computations are now easily completed and similar arguments complete the proof of Statement 8.

The following statement enables us to give a fairly clear picture of the approximate location of those orbits which make a transit of the region near the equilibrium point ( $\alpha_1 \alpha_2 < 0$ ). This picture carries over to the restricted problem with little difficulty and suggests a "possible" means of giving an existence proof for the periodic orbits of M. Davidson. (However, the present author has not been able to carry out any proof as yet.)

Before giving this statement, we state a lemma. In the lemma,  $\cos^{-1}(\gamma)$  denotes that angle between 0 and  $\pi$  whose cosine is  $\gamma$ : (provided  $|\gamma| < 1$ )

Lemma:

$$\alpha \cos \theta + \beta \sin \theta \geq \gamma \iff |\chi - \theta| \leq \cos^{-1} \frac{\gamma}{(\alpha^2 + \beta^2)^{\frac{1}{2}}}$$

where

$$\cos \chi \sim \alpha; \quad \sin \chi \sim \beta$$

the equality signs hold simultaneously. If  $\gamma^2 > \alpha^2 + \beta^2$  the inequality never holds.

Statement 9

Let  $z_1 = \rho \cos \theta$ ;  $z_2 = \rho \sin \theta$ .

Let  $x = (x_1, x_2)$  denote any point in R.

If  $\gamma_1 = -\frac{ax_1 + b\sigma x_2}{\mu\rho}$ ;  $\gamma_2 = \frac{ax_1 - b\sigma x_2}{\mu\rho}$

$$\cos \chi_1 \sim 1$$

$$\cos \chi_2 \sim 1$$

$$\sin \chi_1 \sim -\sigma$$

$$\sin \chi_2 \sim \sigma$$

a) Then for  $|\gamma_2| \leq 1$ , we have:

$$\alpha_i \geq 0 \iff |\theta - \chi_i| \leq \cos^{-1} \frac{\gamma_i}{(1 + \sigma^2)^{\frac{1}{2}}}.$$

b) It follows (Statement 8) that  $|\gamma_1| \leq (1 + \sigma^2)^{\frac{1}{2}}$  only in the strip between the lines enveloping the orbits with  $\alpha_1 = 0$  and that  $|\gamma_1| = (1 + \sigma^2)^{\frac{1}{2}}$  on the boundary of these strips.



### Proof of Statement 9

From Statement 2, we have

$$|e_1| \alpha_1 = a x_1 + b \sigma x_2 + \mu z_1 - \mu \sigma z_2$$

$$|e_1| \alpha_2 = a x_1 - b \sigma x_2 - \mu z_1 - \mu \sigma z_2$$

Replacing  $z_1$  by  $\rho \cos \theta$  and  $z_2$  by  $\rho \sin \theta$  we set

$$\alpha_1 \geq 0 \iff \cos \theta + \sigma \sin \theta \geq - \frac{(a x_1 + b \sigma x_2)}{\mu \rho}$$

and

$$\alpha_2 \geq 0 \iff \cos \theta + \sigma \sin \theta \geq \frac{(a x_1 - b \sigma x_2)}{\mu \rho}$$

An application of the lemma completes the proof.

### Statement 10 (consequence of 9)

From 9, it follows that orbits with  $\alpha_2 = 0$  cut the line  $\gamma_2 = 0$  in a direction orthogonal to the enveloping lines of these orbits ( $i = 1, 2$ ). Thus the lines  $\mu_2 = 0$  must pass through the points of tangency of the enveloping lines with the boundary of  $R$ .

We further observe that to the "right" of the line  $\gamma_i = 0$ ,  $\chi_i$  is acute, while to the left of the line  $\gamma_i = 0$ ,  $\chi_i$  is obtuse. The results implied by figure 1 are easy consequences. In particular, we see for example that any orbits in the regions I, I', I''; II, II', II'' have  $\alpha_1 \alpha_2 > 0$  while those in the regions III, III' have  $\alpha_1 \alpha_2 < 0$ . The situation in the strips is not as simple, but is fairly clear.

### Figure 1.

1) The (two) solid dark lines through the points A and D are the enveloping lines of solutions with  $\alpha_1 = 0$ . The corresponding lines through B and C are the enveloping lines of solutions with  $\alpha_2 = 0$ . Any solution with  $\alpha_1 = 0$  or  $\alpha_2 = 0$  must lie in the corresponding strip bounded by these lines.

2) At P, the shaded wedge indicates the directions at P for which the corresponding solution has  $\alpha_1 \leq 0$ . At P' the shaded wedge indicates  $\alpha_1 \geq 0$ . Similarly at Q the wedge indicates  $\alpha_2 < 0$ , at Q',  $\alpha_2 > 0$ . On the dotted line  $\overline{AD}$  the wedge has angle  $\pi$  corresponding to  $\gamma_1 = 0$ .  $\overline{CB}$  has a similar meaning with regard to the strip for  $\alpha_2$ .

3) The solid lines parallel to the strips indicate the regions where the corresponding  $\alpha_i > 0$  for all possible angles. The dotted lines similarly indicate where  $\alpha_i < 0$ .

4) Thus we can see that in regions I, I', I'', both of  $\alpha_1, \alpha_2$  are positive, while in the regions II, II', II'',  $\alpha_1$  and  $\alpha_2$  are negative. Finally in regions III,  $\alpha_1 > 0$ ;  $\alpha_2 < 0$  while in III',  $\alpha_1 < 0$ ;  $\alpha_2 > 0$ .

5) In the strips we must determine the sign of  $\alpha$  from the direction of the velocity vector: e.g., at P, any solution whose velocity vector lies in the shaded wedge has  $\alpha_2 > 0$ ,  $\alpha_1 < 0$ , etc.

Thus we have a geometric criterion for determining whether or not a solution will make a transit of the equilibrium region. Note in particular that such a solution must stay inside one or the other of the strips away from the equilibrium, and that as it crosses the equilibrium region it changes strips. Solutions going from right to left are "on the bottom"; those from left to right on top.

We conclude with a remark which may have some "engineering" value:

Statement 11. The (two) solutions for which  $|\beta| = 0$  are hyperbolas; these solutions correspond to those orbits which cross the region of the equilibrium point the fastest.

(Corresponding solutions for the restricted problem exist and are well approximated by these — in the equilibrium region — for energies slightly larger than that of the equilibrium.)

The equation for these orbits are

$$-a x_1 = v \tau z_2$$

$$-b \tau x_2 = v z_1$$

or

$$\begin{aligned} -\sigma x_1^2 + x_2^2 &\approx 2h\nu^2 (b\tau^2 + \nu^2)^{-1} b^{-1} \\ &= \frac{2h\nu^2}{e_2 b} \end{aligned}$$

Proof: Statement 8 plus some algebra.

(Note that the left hand side is determined from geometrical considerations alone, while the right hand side follows by letting  $x_1 \approx 0$  and using the energy equation.)

This completes the present collection of statements.

#### BIBLIOGRAPHY

1. Davidson, M. (to be published).
2. Moser, J., On the Generalization of a Theorem of A. Liapounoff, Comm. Pure Appl. Math., Vol. XI, 1958, pp. 257-271.

HAYES INTERNATIONAL CORPORATION

N65 33059

ON AN APPLICATION OF THE HAMILTON-JACOBI  
THEORY TO HIGH THRUST ROCKET FLIGHT

A. A. NAFOOSI  
H. PASSMORE

BIRMINGHAM, ALABAMA

## INTRODUCTION

To date, no closed form solution of the equations representing minimum fuel flight of a high thrust vehicle operating in a vacuum under an inverse square gravitational attraction has been determined. Optimum trajectories, under these conditions, must therefore be calculated by numerical methods and iteration techniques.

On the other hand, the powerful methods of classical (or variational) mechanics hold promise of solving "all" dynamical problems. The "only" difficulty being the establishment of a Hamiltonian function in a separable form. The solution of "all" dynamical problems using these methods will therefore not be imminent pending the development of a general transformation procedure that will transform the Hamiltonian of any given problem into a separable form.

This paper presents a brief discussion of the classical procedures, discusses both closed form and several approximate solution procedures and shows the level of application to the minimum fuel trajectory problem.

## THE PROBLEM

The physical problem will be taken to be the determination of trajectories for minimum fuel consumption for vehicle flight in a vacuum under the influence of a high level, constant thrust and an inverse square gravity field. This of course is not the most general problem which could include variable thrust levels, higher order gravitational attractions, atmospheric loads and disturbances, and numerous other variables. However, it is general enough to describe most of the solution difficulties inherent in this type of problem.

The two dimensional equations of motion of a point mass vehicle subjected to the forces described above may be expressed in cartesian coordinates as

$$\begin{aligned}\ddot{x} &= \frac{F}{m} \sin \chi - \frac{\mu}{r^3} x \\ \ddot{y} &= \frac{F}{m} \cos \chi - \frac{\mu}{r^3} y\end{aligned}\tag{1}$$

where  $x$  and  $y$  are horizontal and vertical coordinates respectively,  $\mu$  is the gravitational constant,  $F$  is the constant thrust,  $m$  is the vehicle mass,  $\chi$  is the angle of thrust direction measured from the vertical and  $r$  is the radius or distance of the vehicle from the center of attraction ( $r = [x^2 + y^2]^{1/2}$ ). Specifying now that the mass flow  $\dot{m}$  shall

be maintained at a constant rate  $K$ , and introducing new variables as:

$$\dot{q}_1 = \dot{x}, \quad \dot{q}_2 = \dot{y}, \quad q_3 = x, \quad q_4 = y, \quad q_5 = m \quad (2)$$

The equations of motion in first order form become

$$\begin{aligned} \dot{q}_1 &= \frac{F}{q_5} \sin \chi - \frac{\mu}{r^3} q_3 \\ \dot{q}_2 &= \frac{F}{q_5} \cos \chi - \frac{\mu}{r^3} q_4 \\ \dot{q}_3 &= q_1 \\ \dot{q}_4 &= q_2 \\ \dot{q}_5 &= -K \end{aligned} \quad (3)$$

It is noted, that due to the constancy restriction on the mass flow, a minimum fuel trajectory is now analogous to a minimum time trajectory. The problem now is to determine the control variable  $\chi$  such as to insure that any trajectory obtained through an integration of equations (3) will be a minimum time (fuel) trajectory. It is therefore necessary to apply some analytical optimization technique. Both the Calculus of Variations and Pontryagin's Maximum Principle are usable here and yield identical results. However, since it will be necessary to have a Hamiltonian available for later applications, the Pontryagin technique (Reference 1) will be used.

Defining the auxiliary variables as  $p_i$  ( $i=1, \dots, 5$ ), the Pontryagin Hamiltonian function becomes

$$H(p_i, q_i, \chi) = \frac{F}{q_5} (p_1 \sin \chi + p_2 \cos \chi) - \frac{\mu}{r^3} (p_1 q_3 + p_2 q_4) + p_3 q_1 + p_4 q_2 - p_5 K \quad (4)$$

The condition that this function maintain a maximum is then

$$\frac{\partial H}{\partial \chi} = 0 = \frac{F}{q_5} (p_1 \cos \chi - p_2 \sin \chi) \quad (5)$$

from which

$$\tan \chi = p_1 / p_2$$

Hence

$$\sin \chi = \frac{p_1}{(p_1^2 + p_2^2)^{1/2}} ; \quad \cos \chi = \frac{p_2}{(p_1^2 + p_2^2)^{1/2}} \quad (6)$$

Substitution of equation (6) in (4) then yields

$$H(p_i, q_i) = \frac{F}{q_5} (p_1^2 + p_2^2)^{1/2} - \frac{\mu}{r^3} (p_1 q_3 + p_2 q_4) + p_3 q_1 + p_4 q_2 - p_5 K \quad (7)$$

The equations may then be expressed:

$$\dot{q}_i = \partial H / \partial p_i ; \quad \dot{p}_i = - \partial H / \partial q_i \quad (8)$$

(i = 1, ---5)

The problem of obtaining the optimum trajectory now becomes the problem of integrating equations (8).



## APPROACH

The approach used to study solutions of equations 6 is the Hamilton-Jacobi theory of canonical transformations. This theory was developed for basic dynamical systems, however it is applicable to any system whose governing equations may be expressed in first order form as

$$\dot{q}_i = \partial F / \partial p_i \quad ; \quad \dot{p}_i = -\partial F / \partial q_i \quad (i = 1 \dots n) \quad (9)$$

Where the function  $F(q_i, p_i)$  is not restricted to the Hamiltonian of classical mechanics, but can be any function which allows presentation in the above canonical form. It is, however, usually referred to as the Hamiltonian function or simply the Hamiltonian.

Now, examining the equations (9), it is seen that if one of the  $q_i$  (or  $p_i$ ) is not present in the Hamiltonian (i. e. if a variable is cyclic or ignorable) then the partial derivative of  $F$  with respect to that variable is zero and the corresponding  $p_i$  (or  $q_i$ ) is constant. Consequently, if the system can be transformed to a new system of coordinates, while maintaining the canonical form, such that all of the new coordinates and their conjugates except one is cyclic, then the problem is solved. The most direct way to do this is to set the Hamiltonian itself equal to the one non cyclic new coordinate.

$$F'(P_i, Q_i) = Q_i$$

which gives

$$\begin{aligned}\dot{Q}_1 &= 0 & \dot{P}_1 &= -1 \\ \dot{Q}_i &= 0 & \dot{P}_i &= 0 \quad (i = 2, \dots, n)\end{aligned}\tag{10}$$

Hence all  $Q_i = \text{constant} = \alpha_i$  and all  $P_i = \text{constant} = \beta_i$  except  $P_1 = (\beta_1 - t)$ .

It is however necessary to determine the canonical coordinate transformation required to transform

$$F(q_i, p_i) \Rightarrow F'(Q_i, P_i) = Q_1\tag{11}$$

To do this, it is necessary to introduce a generating function \*,

$S(q_i, P_i)$  a function of one set of old variables and one set of new variables. The transformation equations may then be written

$$p_i = \partial S / \partial q_i \quad ; \quad Q_i = \partial S / \partial P_i \quad (i = 1 \dots n)\tag{12}$$

$S(q_i, P_i)$ , however must still be determined. This may be done (theoretically at least) by substituting the applicable transformation equation

$$p_i = \partial S / \partial q_i$$

into the old Hamiltonian and setting it equal to the new Hamiltonian

$$F(q_i, \partial S / \partial q_i) = Q_1\tag{13}$$

\*There will be no discussion here as to the basic differences between Hamilton's Principle function  $W$  and Jacobi's function  $S$ . Also  $S$  may take any of the four forms  $S(q_i, P_i)$ ,  $S(q_i, Q_i)$ ,  $S(Q_i, p_i)$  or  $S(p_i, P_i)$  as needed in a particular problem. A discussion of these areas appears in Reference 2.

$S(q_i, P_i)$  is then determined through the solution of the partial differential equation (13) which is usually referred to as the Hamilton-Jacobi equation. With  $S(q_i, P_i)$  known, the necessary transform relations may be obtained from equations (12).

$$\begin{aligned} p_j &= p_j(Q_i, P_i) \\ q_j &= q_j(Q_i, P_i) \end{aligned} \quad j(1 \dots n) \quad (14)$$

One further item, the canonical perturbation technique, might be mentioned before concluding this discussion of the procedure used. Often, it is possible to divide the Hamiltonian into the sum of two parts one of which may be considered as a perturbation. The equations then appear as

$$q_i = \frac{\partial F_o}{\partial p_i} - \frac{\partial F_1}{\partial p_i} ; \quad p_i = \frac{-\partial F_o}{\partial q_i} + \frac{\partial F_1}{\partial q_i} \quad (15)$$

where  $F = F_o - F_1$

The procedure then is to neglect the  $F_1$  portion and solve the equations

$$\dot{q}_i = \frac{\partial F_o}{\partial p_i} ; \quad \dot{p}_i = - \frac{\partial F_o}{\partial q_i} \quad (16)$$

using a generating function  $S(q, P)$  and the Hamilton-Jacobi relations to obtain

$$q_j = q_j(Q_i, P_i) ; \quad p_j = p_j(Q_i, P_i) \quad (17)$$

These solutions (17) to the first part of the problem are then substituted into the original  $F = F_o - F_1$  and into the original equations (15).

Then, after several straightforward, but lengthy, manipulations (See References 3 and 4), the following equations in the new variables result.

$$\dot{Q}_i = \frac{\partial F_1(P_i, Q_i)}{\partial P_i} ; \quad \dot{P}_i = \frac{-\partial F_1(P_i, Q_i)}{\partial Q_i} \quad (18)$$

Hopefully then,  $F_1(P_i, Q_i)$  is in a simple form such that the Hamilton-Jacobi equation for this part of the problem may be solved either completely or approximately.

The net result of these procedures, whether the direct approach or a perturbation technique is used, is that the problem of integrating the original equations of motion, equations (9), has been "reduced" to the problem of finding a solution of the Hamilton-Jacobi equation, equation (13).

## SOLUTION OF HAMILTON-JACOBI EQUATIONS

Two methods which give closed form solutions to some Hamilton-Jacobi equations are the method of separation of variables (Reference 2) and a closely related though more orderly method known as Jacobi's Method (Reference 5). The method of separation of variables is probably the easiest method of solving the Hamilton-Jacobi equation when it is applicable. However, in application the method is not well organized and is quite dependent upon the skill of the operator to "see" the separation. Also, the question of whether or not the equation is separable depends upon the coordinates employed. The restricted two body problem is separable in polar (or spherical) coordinates, but not in cartesian, and the coordinates for which the famous three body problem is separable have evaded investigators for years.

Some insight into whether or not the H-J equation is separable in a particular system of coordinates may be gained through the development of a separation criteria.

The real question of separability is the question of whether functions of the form

$$p_i = p_i(q_i, \alpha_1, \dots, \alpha_n) \quad (19)$$

can be found so that when substituted in

$$H(q_1, q_2, \dots, q_n, p_1, p_2, \dots, p_n) = E \quad (20)$$

will cancel out all the  $q_i$ 's is to be answered.

Our purpose is to find the condition that the Hamiltonian be separable with respect to a set of coordinates. If this condition is not satisfied in one set of coordinates, then one needs to find the proper coordinates which satisfy the condition.

Now let us assume that we can find  $p_i$  as in (19) which satisfies (20). It follows that  $p_i$  and its derivative with respect to  $q_i$  are functions of a single coordinate  $q_i$ . Differentiate (20) with respect to  $q_i$ , we obtain:

$$\frac{\partial H}{\partial q_i} + \frac{\partial H}{\partial p_i} \frac{\partial p_i}{\partial q_i} = 0 \quad (21)$$

Let us introduce a new function  $\rho_i$  of the form:

$$\rho_i = f(q_1, q_2, \dots, q_n; p_1, p_2, \dots, p_n) \quad (22)$$

such that it will satisfy the relation:

$$\frac{\partial H}{\partial q_i} - \frac{\partial H}{\partial p_i} \rho_i = 0 \quad (23)$$

By comparing (19) and (21) we obtain:

$$\rho_i = - \frac{\partial p_i}{\partial q_i} \quad (23a)$$

and thus  $\rho_i$  is a function of  $q_i$  alone, since  $p_i$  is a function of  $q_i$  alone by (19). By differentiating (23) with respect to  $q_j$ , and keeping in mind relation (22), we obtain

$$\frac{\partial \rho_i}{\partial q_j} + \frac{\partial \rho_i}{\partial p_j} \frac{\partial p_j}{\partial q_j} = 0 \quad \text{for } j \neq i \quad (23b)$$

From (23a) and (23b), we obtain:

$$\frac{\partial \rho_i}{\partial q_j} - \frac{\partial \rho_i}{\partial P_j} \rho_j = 0 \quad (23c)$$

Differentiating equation (23) with respect to  $q_j$ , and using (23a), we obtain:

$$\frac{\partial^2 H}{\partial q_i \partial q_j} - \frac{\partial^2 H}{\partial q_i \partial P_j} \rho_j - \left( \frac{\partial^2 H}{\partial P_i \partial q_j} - \frac{\partial^2 H}{\partial P_i \partial P_j} \rho_j \right) \rho_i - \frac{\partial H}{\partial P_i} \left( \frac{\partial \rho_i}{\partial q_j} \right) = 0$$

Using (21) and  $\frac{\partial \rho_i}{\partial q_j} = 0$ , we obtain:

$$\frac{\partial^2 H}{\partial q_i \partial q_j} - \frac{\partial^2 H}{\partial q_i \partial P_j} \frac{\partial H / \partial q_j}{\partial H / \partial P_j} - \left( \frac{\partial^2 H}{\partial P_i \partial q_j} - \frac{\partial^2 H}{\partial P_i \partial P_j} \frac{\partial H / \partial q_j}{\partial H / \partial P_j} \right) \frac{\partial H / \partial q_i}{\partial H / \partial P_i} = 0$$

By simplification

$$\frac{\partial^2 H}{\partial q_i \partial q_j} \frac{\partial H}{\partial P_i} \frac{\partial H}{\partial P_j} - \frac{\partial^2 H}{\partial q_i \partial P_j} \frac{\partial H}{\partial q_j} \frac{\partial H}{\partial P_i} - \frac{\partial^2 H}{\partial P_i \partial q_j} \frac{\partial H}{\partial q_i} \frac{\partial H}{\partial P_j} + \frac{\partial^2 H}{\partial P_i \partial P_j} \frac{\partial H}{\partial q_i} \frac{\partial H}{\partial q_j} = 0 \quad (24)$$

for  $i, j = 1, 2, \dots, n$  and  $i \neq j$

Therefore, the necessary condition that (20) be separable is condition (24).

It can be easily shown that the validity of equation (24) is also sufficient for the integration through separation of variables.

One interesting case of separability is the case where the motion is known to be periodic. In this case, the proper coordinates are the action and the angle variables, and the H-j equation is separable.

If the H-j equation is separable in more than one set of coordinates, then this case is said to be degenerate. There is similarity between

this degeneracy and the general one. Consider the general equation:

$$QX = Q_0 X$$

where  $Q$  is an operator,  $Q_0$  is a constant, and  $X$  is the characteristic function. It is clear that for each value of  $Q_0$  there corresponds one or more  $X$ . In case there is only one  $X$ , then  $Q$  is said to be non-degenerate, otherwise it is called degenerate.

The similarity of the H-j equation with the general case above can be visualized by taking the Hamiltonian,  $H$ , as the operator,  $Q$ , the constant,  $\alpha$ , as  $Q_0$  and the generating function,  $S$  as  $X$ . If we define the Hamiltonian operator:  $H = H(\alpha_i, X_i, \frac{\partial}{\partial X_i})$

as having the property

$$H(\alpha_i, X_i, \frac{\partial}{\partial X_i}) S = H(\alpha_i, X_i, \frac{\partial S}{\partial X_i})$$

$$H(\alpha_i, X_i) S = S$$

then our H-j equation will take the form:

$$HS = \alpha S$$

Thus  $H$  is degenerate or non-degenerate according to the number of solutions of  $S$  if it is one or more. This is equivalent to saying that the equation is separable in one set of coordinates or more.

As mentioned before, "Jacobi's" method for obtaining solutions to first order partial differential equations appears more orderly



than the separation techniques. Since this method does not appear to frequent the literature as much as the separation procedures, a brief development is presented here.

The solution of the H-j equation involves the determination of the generating function, S. The Hamilton-Jacobi equation may be written in the form

$$F(q_1, q_2, \dots, q_n, P_1, P_2, \dots, P_n) = 0 \quad (25)$$

where

$$F = H - \alpha; \quad P_i = \frac{\partial S}{\partial q_i} \quad \text{and } H \text{ is the Hamiltonian.}$$

Second, we try to find (n-1) compatible functions to F, i.e. (n-1)

additional functions  $F_i$ 's, which satisfy (25), i.e.,

$$F_i(q_1, q_2, \dots, q_n; P_1, P_2, \dots, P_n) = \alpha_i, \quad (i = 1, 2, \dots, n-1) \quad (26)$$

where the  $\alpha_i$  are arbitrary constants. Third, the  $P_1, P_2, \dots, P_n$  can be determined from (25) and (26) as functions of q's and  $\alpha$ 's and such that these functions, when inserted in the differential relation

$$dS = P_1 dq_1 + P_2 dq_2 + \dots + P_n dq_n \quad (27)$$

yield an integrable equation. The result of integrating (27) whereby an arbitrary constant  $\alpha_n$  is introduced, is our generating function.

Since the proof is too long and complicated in the general case, let us show the procedure for:  $n = 3$ .

$$F(q_1, q_2, q_3, P_1, P_2, P_3) = 0 \quad (28)$$

Let us find two particular integrals of (28) as follows:

$$F_1(q_1, q_2, q_3, P_1, P_2, P_3) = \alpha_1 \quad (29)$$

$$F_2(q_1, q_2, q_3, P_1, P_2, P_3) = \alpha_2 \quad (30)$$

where  $P_1, P_2, P_3$  are functions of  $q_1, q_2, q_3$ .

Since  $F_1, F_2$  are integrals, the "Poisson brackets"

$$[F, F_1] = 0 \quad (31)$$

and

$$[F, F_2] = 0 \quad (32)$$

Moreover,  $F_1$  and  $F_2$  must be compatible, hence

$$[F_1, F_2] = 0 \quad (33)$$

Now solve (28), (29), and (30) for  $P_1, P_2, P_3$  and form

$$dS = P_1 dq_1 + P_2 dq_2 + P_3 dq_3 \quad (34)$$

which is required to be integrable.

In order to find the relations between the  $F_i$ 's and  $P_i$ 's which

satisfy the above conditions, we expand (31) in the usual form:

$$\frac{\partial F}{\partial q_1} \frac{\partial F_1}{\partial P_1} + \frac{\partial F}{\partial q_2} \frac{\partial F_1}{\partial P_2} + \frac{\partial F}{\partial q_3} \frac{\partial F_1}{\partial P_3} - \frac{\partial F}{\partial P_1} \frac{\partial F_1}{\partial q_1} - \frac{\partial F}{\partial P_2} \frac{\partial F_1}{\partial q_2} - \frac{\partial F}{\partial P_3} \frac{\partial F_1}{\partial q_3} = 0$$

This is a homogeneous linear partial (differential equation) for de-

termining  $F_1$ . Its subsidiary equations are

$$\frac{\frac{dP_1}{\partial F}}{\frac{\partial q_1}{\partial F}} = \frac{\frac{dP_2}{\partial F}}{\frac{\partial q_2}{\partial F}} = \frac{\frac{dP_3}{\partial F}}{\frac{\partial q_3}{\partial F}} = \frac{\frac{dq_1}{\partial F}}{\frac{\partial P_1}{\partial F}} = \frac{\frac{dq_2}{\partial F}}{\frac{\partial P_2}{\partial F}} = \frac{\frac{dq_3}{\partial F}}{\frac{\partial P_3}{\partial F}} \quad (35)$$

These relations also serve as subsidiary equation to (32) for the determination of  $F_2$ . Therefore, if one finds from (35) two independent integrals  $F_1 = \alpha_1$  and  $F_2 = \alpha_2$ , then all the relations (31), (32) and (33) will be fulfilled, and our task is accomplished. The procedure for the general case is exactly the same.

If given the partial differential equation

$$F(q_1, q_2, \dots, q_n; P_1, P_2, \dots, P_n) = 0$$

then, form the subsidiary equations

$$\frac{\frac{dP_1}{\partial F}}{\partial q_1} = \frac{\frac{dP_2}{\partial F}}{\partial q_2} = \dots = \frac{\frac{dP_n}{\partial F}}{\partial q_n} = \frac{\frac{dq_1}{\partial F}}{\partial P_1} = \frac{\frac{dq_2}{\partial F}}{\partial P_2} = \dots = \frac{\frac{dq_n}{\partial F}}{\partial P_n}$$

and find  $(n-1)$  independent integrals

$$F_i = \alpha_i \quad i = 1, 2, \dots, n-1$$

such that

$$[F_i, F_j] = 0 \quad i, j = 1, 2, \dots, n-1 \quad i \neq j$$

Then solve the  $n$  equations

$$F = 0 \text{ and } F_i = \alpha_i \quad i = 1, 2, \dots, n-1$$

for the  $P$ 's in terms of  $q$ 's and  $\alpha$ 's, and insert their expressions in

$$dS = P_1 dq_1 + P_2 dq_2 + \dots + P_n dq_n$$

Integration of this equation leads to a complete integral of  $S$ .

## APPLICATION TO THE PROBLEM

Canonical perturbation techniques (using Jacobi's method to solve the Hamilton-Jacobi equation) may be applied to the problem of equations (6). To illustrate, consider equations (6)

$$\dot{q}_i = \frac{\partial H}{\partial p_i} \quad ; \quad \dot{p}_i = -\partial H / \partial q_i \quad (i = 1, \dots, 5) \quad (6)$$

with H given by equation (5) as

$$H = \frac{F}{q_5} \sqrt{p_1^2 + p_2^2} + p_3 q_1 + p_4 q_2 - p_5 K - \frac{GM}{r^3} (p_1 q_3 + p_2 q_4) \quad (5)$$

define

$$H_0 = \frac{F}{q_5} \sqrt{p_1^2 + p_2^2} + p_3 q_1 + p_4 q_2 - p_5 K \quad (36a)$$

and

$$H_1 = \frac{GM}{r^3} (p_1 q_3 + p_2 q_4) \quad (36b)$$

The equations are then expressed

$$\dot{q}_i = \frac{\partial H_0}{\partial p_i} - \frac{\partial H_1}{\partial p_i} \quad ; \quad \dot{p}_i = -\frac{\partial H_0}{\partial q_i} + \frac{\partial H_1}{\partial q_i} \quad (i = 1 \dots 5) \quad (37)$$

### ZERO GRAVITY APPROXIMATION

Consider first the problem

$$\dot{q}_i = \frac{\partial H_0}{\partial p_i} \quad ; \quad \dot{p}_i = -\frac{\partial H_0}{\partial q_i} \quad (38)$$

The Hamilton - Jacobi equation for this problem is

$$\frac{F}{m} \sqrt{p_1^2 + p_2^2} + p_3 q_1 + p_4 q_2 - p_5 K - P_5 = 0 \quad (39)$$

Where  $P_5$  is the introduced constant. The subsidiary equations (analogous to equations 35) of Jacobi's method are then

$$\begin{aligned} \frac{dp_1}{p_3} &= \frac{dp_2}{p_4} = \frac{dp_3}{0} = \frac{dp_4}{0} = \frac{dp_5}{-\frac{F}{q_5} \sqrt{p_1^2 + p_2^2}} \\ &= \frac{dq_1}{-\frac{E}{q_5} \frac{p_1}{\sqrt{p_1^2 + p_2^2}}} = \frac{dq_2}{-\frac{F}{q_5} \frac{p_2}{\sqrt{p_1^2 + p_2^2}}} = \frac{dq_3}{-q_1} = \frac{dq_4}{-q_2} = \frac{dq_5}{K} \end{aligned} \quad (40)$$

The third and fourth conditions give

$$p_3 = P_3 = \text{const.} \quad (41)$$

$$p_4 = P_4 = \text{const.} \quad (42)$$

as expected since  $p_3$  and  $p_4$  are cyclic in  $H_0$ .

From the first and last of equations (40):

$$\begin{aligned} dp_1 &= \frac{P_3}{K} dq_5 \\ p_1 &= \frac{P_3}{K} q_5 + P_1 \end{aligned} \quad (43)$$

From the second and last of equations (40)

$$\begin{aligned} dp_2 &= \frac{P_4}{K} dq_5 \\ p_2 &= \frac{P_4}{K} q_5 + P_2 \end{aligned} \quad (44)$$

Then, substituting equations (41), (42), (43), and (44) into equation

(39),  $p_5$  becomes

$$\begin{aligned} p_5 &= \frac{F}{q_5} K \left[ \left( \frac{P_3}{K} q_5 + P_1 \right)^2 + \left( \frac{P_4}{K} q_5 + P_2 \right)^2 \right]^{1/2} \\ &\quad + \frac{P_3}{K} q_1 + \frac{P_4}{K} q_2 - \frac{P_5}{K} \end{aligned} \quad (45)$$

The equation for obtaining the generating function (analogous to equation 34) is

$$dS = p_1 dq_1 + p_2 dq_2 + p_3 dq_3 + p_4 dq_4 + p_5 dq_5$$

Substitution for the  $p_i$  and integrating gives:

$$S = \left( \frac{P_3}{K} q_5 + P_1 \right) q_1 + \left( \frac{P_4}{K} q_5 + P_2 \right) q_2 + P_3 q_3 + P_4 q_4 - \frac{P_5}{K} q_5 + \frac{F}{K} \left\{ C + \frac{P_1 P_3 + P_2 P_4}{\sqrt{P_3^2 + P_4^2}} \ln A - \sqrt{P_1^2 + P_2^2} \ln B \right\} \quad (46)$$

$$\text{where } C = \sqrt{\left( \frac{P_3}{K} q_5 + P_1 \right)^2 + \left( \frac{P_4}{K} q_5 + P_2 \right)^2} \quad (47a)$$

$$A = \frac{2}{K} \left[ \sqrt{P_3^2 + P_4^2} C + (P_3^2 + P_4^2) \frac{q_5}{K} + (P_1 P_3 + P_2 P_4) \right] \quad (47b)$$

$$B = \frac{2}{q_5} \left[ \sqrt{P_1^2 + P_2^2} C + (P_1^2 + P_2^2) + (P_1 P_3 + P_2 P_4) \frac{q_5}{K} \right] \quad (47c)$$

The transform relations are then obtained from

$$Q_i = \partial S / \partial P_i$$

as

$$q_1 = Q_1 - \frac{F}{K} \left\{ \frac{(P_1 - P_3 Q_5)}{C} + \frac{P_3}{\sqrt{P_3^2 + P_4^2}} \ln A - \frac{P_1}{\sqrt{P_1^2 + P_2^2}} \ln B + \frac{2(P_1 P_3 + P_2 P_4)}{AK \sqrt{P_3^2 + P_4^2}} \left[ \frac{\sqrt{(P_3^2 + P_4^2)}(P_1 - P_3 Q_5)}{C} + P_3 \right] + \frac{2\sqrt{P_1^2 + P_2^2}}{BK Q_5} \left[ \frac{P_1 C}{\sqrt{P_1^2 + P_2^2}} + \frac{\sqrt{P_1^2 + P_2^2} (P_1 - P_3 Q_5)}{C} + 2P_1 - P_3 Q_5 \right] \right\} \quad (48)$$

$$q_2 = Q_2 - \frac{F}{K} \left\{ \frac{(P_2 - P_4 Q_5)}{C} + \frac{P_4}{\sqrt{P_3^2 + P_4^2}} \ln A - \frac{P_2}{\sqrt{P_1^2 + P_2^2}} \ln B + \frac{2(P_2 P_3 + P_2 P_4)}{AK \sqrt{P_3^2 + P_4^2}} \left[ \frac{\sqrt{P_3^2 + P_4^2} (P_2 - P_4 Q_5)}{C} + P_4 \right] + \frac{2\sqrt{P_1^2 + P_2^2}}{BK Q_5} \left[ \frac{\sqrt{P_1^2 + P_2^2} (P_2 - P_4 Q_5)}{C} + \frac{P_2 C}{\sqrt{P_1^2 + P_2^2}} + 2P_2 - P_4 Q_5 \right] \right\} \quad (49)$$

$$\begin{aligned}
q_3 = & Q_3 + Q_1 Q_5 - \frac{F}{K} \left\{ \frac{2(P_1 P_3 + P_2 P_4)}{AK \sqrt{P_3^2 + P_4^2}} \left[ P_1 - P_3 Q_5 + \frac{P_3 C}{\sqrt{P_3^2 + P_4^2}} \right] \right. \\
& + \frac{2\sqrt{P_1^2 + P_2^2}}{BK} \left[ P_1 - P_3 Q_5 + \frac{P_1 C}{\sqrt{P_1^2 + P_2^2}} \right] + \frac{1}{\sqrt{P_3^2 + P_4^2}} \left[ P_1 + P_3 Q_5 \right. \\
& - \left. \frac{P_3(P_1 P_3 + P_2 P_4)}{(P_3^2 + P_4^2)} \right] \ln A \\
& - \left. \frac{P_1 Q_5}{\sqrt{P_1^2 + P_2^2}} \ln B \right\} \quad (50)
\end{aligned}$$

$$\begin{aligned}
q_4 = & Q_4 + Q_2 Q_5 - \frac{F}{K} \left\{ \frac{2(P_1 P_3 + P_2 P_4)}{AK \sqrt{P_3^2 + P_4^2}} \left[ P_2 - P_4 Q_5 + \frac{P_4 C}{\sqrt{P_3^2 + P_4^2}} \right] \right. \\
& + \frac{2\sqrt{P_1^2 + P_2^2}}{BK} \left[ P_2 - P_4 Q_5 + \frac{P_2 C}{\sqrt{P_1^2 + P_2^2}} \right] \quad (51) \\
& + \frac{1}{\sqrt{P_3^2 + P_4^2}} \left[ P_2 + P_4 Q_5 - \frac{(P_1 P_3 + P_2 P_4) P_4}{(P_3^2 + P_4^2)} \right] \ln A - \frac{P_2 Q_5}{\sqrt{P_1^2 + P_2^2}} \ln B \Big\}
\end{aligned}$$

$$q_5 = -KQ_5 \quad (52)$$

#### CONSTANT GRAVITY - FLAT EARTH

It is now desired to perturb this zero gravity solution into a solution to the constant gravity flat earth problem. The equations are

$$\text{then} \quad \dot{Q}_i = \frac{\partial H'}{\partial P_i} \quad \dot{P}_i = -\frac{\partial H'}{\partial Q_i} \quad (53)$$

where

$$H' = P_5 - g(P_2 - P_4 Q_5) \quad (54)$$

Specifying a determining function  $W = W(Q_i, \lambda_i)$ , the Hamilton-

Jacobi equation becomes

$$\frac{\partial W}{\partial Q_5} - g \frac{\partial W}{\partial Q_2} + g Q_5 \frac{\partial W}{\partial Q_4} - \lambda_5 = 0 \quad (55)$$

Assuming a solution for W as

$$W = W_1(Q_1) + W_2(Q_2) + W_3(Q_3) + W_4(Q_4) + W_5(Q_5) \quad (56)$$

The Hamilton-Jacobi equation becomes

$$\frac{\partial W_5}{\partial Q_5} + g Q_5 \frac{\partial W_4}{\partial Q_4} - g \frac{\partial W_2}{\partial Q_2} - \lambda_5 = 0 \quad (57)$$

Since the coordinates  $Q_1, Q_2, Q_3$  and  $Q_4$ , are cyclic,  $P_1, P_2, P_3$  and  $P_4$  are constants

$$P_1 = \lambda_1; P_2 = \lambda_2; P_3 = \lambda_3; P_4 = \lambda_4 \quad (58)$$

Hence equation (57) becomes

$$\frac{\partial W_5}{\partial Q_5} + g \lambda_4 Q_5 - (g \lambda_2 + \lambda_5) = 0 \quad (59)$$

which integrates to give

$$W_5 = -\frac{1}{2} g \lambda_4 Q_5^2 + (g \lambda_2 + \lambda_5) Q_5 \quad (60)$$

$W_1$  through  $W_4$  are determined from equations (58) in the form

$$W_j = \lambda_j Q_j \quad (j = 1, 2, 3, 4) \quad (61)$$

Then, W becomes

$$W = \lambda_1 Q_1 + \lambda_2 Q_2 + \lambda_3 Q_3 + \lambda_4 Q_4 + (g \lambda_2 + \lambda_5) Q_5 - \frac{g}{2} \lambda_4 Q_5^2 \quad (62)$$

The coordinates are then obtained from

$$x_i = \partial W / \partial \lambda_i \quad (i = 1 \dots 5)$$

which gives



$$\begin{aligned}
x_1 &= Q_1 \\
x_2 &= Q_2 + gQ_5 \\
x_3 &= Q_3 \\
x_4 &= Q_4 - g/2 Q_5^2 \\
x_5 &= Q_5
\end{aligned} \tag{63}$$

and from equation (53)

$$\lambda_5 = P_5 + gP_4Q_5 - gP_2 \tag{64}$$

and the new Hamiltonian becomes

$$H = \lambda_5$$

with the equations

$$\dot{x}_i = \partial H / \partial \lambda_i \quad ; \quad \dot{\lambda}_i = -\partial H / \partial x_i$$

which gives

$$\begin{aligned}
x_i &= b_i = \text{const} \quad i = 1, 2, 3, 4 \\
\lambda_i &= c_i = \text{const} \quad i = 1, \dots, 5 \\
x_5 &= t + b_5
\end{aligned} \tag{65}$$

Then from 63, 64, 65 and 58

$$\begin{aligned}
Q_1 &= b_1 & P_1 &= c_1 \\
Q_2 &= b_2 - g(b_5 + t) & P_2 &= c_2 \\
Q_3 &= b_3 & P_3 &= c_3 \\
Q_4 &= b_4 - \frac{g}{2} (b_5 + t)^2 & P_4 &= c_4 \\
Q_5 &= (b_5 + t) & P_5 &= c_5 + g c_2 - g c_4 (b_5 + t)
\end{aligned} \tag{66}$$

Then from equations 43, 44, 52, and 66

$$p_1 = c_1 - c_3 (b_5 + t)$$

$$p_2 = c_2 - (b_5 + t)$$

Such that the guidance angle expression becomes

$$\tan \chi = k_0 \left( \frac{k_1 + t}{k_2 + t} \right) \quad (67)$$

where

$$k_0 = c_3 / c_4 ; k_1 = b_5 - c_1 / c_3 ; k_2 = b_5 - c_2 / c_4$$

and  $\tan \chi$  is a bilinear function of time as expected.

#### FORMAT FOR INVERSE SQUARE GRAVITY PERTURBATION

Returning now to the zero gravity solution of equations 41, 42, 43, 44, 45, 48, 49, 50, 51 and 52. Substitution into equation 36 yields the Hamiltonian for the inverse square perturbation term as

$$\begin{aligned} H^* = & P_5 - \frac{GM}{r^3} \left\{ (P_1 - P_3 Q_5) (Q_3 - Q_1 Q_5) + (P_2 - P_4 Q_5) (Q_4 + Q_2 Q_5) \right. \\ & - \frac{F}{K} \left[ \sqrt{(P_1 - P_3 Q_5)^2 + (P_2 - P_4 Q_5)^2} \left( \frac{P_1 P_3 + P_2 P_4}{P_3^2 + P_4^2} - Q_5 \right) \right. \\ & \left. \left. + \frac{1}{\sqrt{P_3^2 + P_4^2}} \left( \frac{(P_1 P_4 - P_2 P_3)}{P_3^2 + P_4^2} - (P_3^2 + P_4^2) Q_5^2 + (P_1 P_3 + P_2 P_4) Q_5 \right) \right] \right. \\ & \left. - \frac{(\ln A)}{\sqrt{P_1^2 + P_2^2}} (P_1^2 + P_2^2 - (P_1 P_3 + P_2 P_4) Q_5) \ln B \right\} \quad (68) \end{aligned}$$

Where A and B are defined in equations 47.

The accompanying equations of motion are then

$$\dot{Q}_i = \partial H^*/\partial P_i, \quad \dot{P}_i = -\partial H^*/\partial Q_i \quad (69)$$

The presence of the numerous radicals and logarithmic terms make the attainment of a solution of the accompanying Hamilton-Jacobi equation quite improbable by ordinary means. Thus, the use of this perturbation method and the Hamilton-Jacobi technique displays little overall advantage in obtaining a closed form solution to the general problem.

#### AN APPROXIMATE SOLUTION - INVERSE SQUARE GRAVITY

The difficulty of obtaining a closed form solution leads to the development of an approximate solution which is taken as a first order improvement on the constant gravity-flat earth solution. Taking the complete Hamiltonian of equation 5, the Hamilton-Jacobi equation may be written

$$\frac{F}{q_5} \sqrt{p_1^2 + p_2^2} + p_3 q_1 + p_4 q_2 - p_5 K - \frac{GM}{r^3} (p_1 q_3 + p_2 q_4) - P_5 = 0 \quad (70)$$

where

$$p_i = \partial S / \partial q_i$$

The subsidiary equations of Jacobi's method are then

$$\begin{aligned}
\frac{dp_1}{p_3} &= \frac{dp_2}{p_4} = -\frac{GM}{r^3} \left[ p_1 - \frac{3q_3}{r^2} (p_1 q_3 + p_2 q_4) \right] = -\frac{GM}{r^3} \left[ p_2 - \frac{3q_4}{r^2} (p_1 q_3 + p_2 q_4) \right] \\
&= -\frac{F}{q_5^2} \frac{dp_5}{\sqrt{p_1^2 + p_2^2}} = -\frac{F}{q_5} \frac{p_1}{\sqrt{p_1^2 + p_2^2}} - \frac{GMq_3}{r^3} = -\frac{F}{q_5} \frac{p_2}{\sqrt{p_1^2 + p_2^2}} - \frac{GMq_4}{r^3} \\
&= \frac{dq_3}{-q_1} = \frac{dq_4}{-q_2} = \frac{dq_5}{K} \quad (71)
\end{aligned}$$

By comparing the above equations with the subsidiary equations of the flat earth problem it is seen that the primary differences are in the denominators of the  $dp_3$  and  $dp_4$  terms and there is an additional term in the  $dq_1$  and  $dq_2$  denominators. Therefore, let the change in the  $p_3$  and  $p_4$  terms be of order  $\epsilon m$  over the constant result of the flat earth problem.

$$\begin{aligned}
p_3 &= P_3 + 2\epsilon_1 q_5 \\
p_4 &= P_4 + 2\epsilon_2 q_5
\end{aligned} \quad (72)$$

where  $\epsilon_1$  and  $\epsilon_2$  are unknown small constants. Substitution into the subsidiary equations then gives: from first and last equation

$$\begin{aligned}
\frac{dp_1}{P_3 + 2\epsilon_1 q_5} &= \frac{dq_5}{K} \\
p_1 &= P_1 + \frac{1}{K} (P_3 q_5 + \epsilon_1 q_5^2) \quad (73)
\end{aligned}$$

similarly from second and last equation

$$\frac{dp_2}{P_4 + 2\epsilon_2 q_5} = \frac{dq_5}{K}$$

$$p_2 = P_2 + \frac{1}{K} (P_4 q_5 + 2 q_5^2) \quad (74)$$

Substitution of 72, 73 and 74 into 70 and solving for  $p_5$  gives

$$p_5 = \frac{1}{K} \left\{ \left[ \left( P_1 + \frac{1}{K} (P_3 q_5 + \epsilon_1 q_5^2) \right)^2 + \left( P_2 + \frac{1}{K} (P_4 q_5 + \epsilon_2 q_5^2) \right)^2 \right]^{1/2} \frac{F}{q_5} \right. \\ \left. + (P_3 + 2\epsilon_1 q_5) q_1 + (P_4 + 2\epsilon_2 q_5) q_2 \right. \\ \left. - \frac{GM}{r^3} \left[ \left( P_1 + \frac{1}{K} (P_3 q_5 + \epsilon_1 q_5^2) \right) q_3 + \left( P_2 + \frac{1}{K} (P_4 q_5 + \epsilon_2 q_5^2) \right) q_4 \right] \right. \\ \left. = P_5 \right\} \quad (75)$$

Now, since  $p_3$  and  $p_4$  were approximated it should not be expected that the  $p$ 's will make the function

$$dS = p_1 dq_1 + p_2 dq_2 + p_3 dq_3 + p_4 dq_4 + p_5 dq_5 \quad (76)$$

an exact differential. Therefore further adjustment must be made

in  $p_5$  to make  $dS$  exact. Hence, assume

$$p_5 = \left\{ \frac{F}{mK} \left[ (P_1^2 + P_2^2) + \frac{2}{K} (P_3 P_1 + P_4 P_2) q_5 + \frac{2}{K^2} \right. \right. \\ \left. \left. (P_3^2 + P_4^2 + 2K\epsilon_1 P_1 + 2K\epsilon_2 P_2) q_5^2 \right]^{1/2} \right. \\ \left. + \frac{1}{K} (P_3 + 2\epsilon_1 q_5) q_1 + \frac{1}{K} (P_4 + 2\epsilon_2 q_5) q_2 + 2\epsilon_1 q_3 + 2\epsilon_2 q_4 - \frac{1}{K} P_5 \right\} \quad (77)$$

Substitution of equations 72, 73, 74 and 77 into equation (76) and integrating then gives

$$\begin{aligned}
 S = & \frac{1}{K} (P_3 q_5 + \epsilon_1 q_5^2) q_1 + P_1 q_1 + \frac{1}{K} (P_4 q_5 + \epsilon_2 q_5^2) q_2 + P_2 q_2 \\
 & + (2\epsilon_1 q_5 + P_3) q_3 + (2\epsilon_2 q_5 + P_4) q_4 - \frac{P_5}{K} q_5 \\
 & + \frac{F}{K} \left\{ C' + \frac{P_3 P_1 + P_2 P_4}{\sqrt{P_3^2 + P_4^2 + 2K\epsilon_1 P_1 + 2K\epsilon_2 P_2}} \ln A' \right. \\
 & \left. - \sqrt{P_1^2 + P_2^2} \ln B' \right\}
 \end{aligned} \tag{78}$$

$$\begin{aligned}
 \text{where } C' = & \frac{\sqrt{(P_1^2 + P_2^2) + \frac{2}{K} (P_3 P_1 + P_4 P_2) q_5 + \frac{2}{K^2} (P_3^2 + P_4^2 + 2K\epsilon_1 P_1 + 2K\epsilon_2 P_2) q_5^2}}{2\sqrt{P_1^2 + P_2^2} C'} \\
 A' = & \frac{2}{K} (P_1 P_3 + P_2 P_4) + \frac{2}{K^2} (P_3^2 + P_4^2 + 2K\epsilon_1 P_1 + 2K\epsilon_2 P_2) q_5 \\
 & + 2\sqrt{P_1^2 + P_2^2} C' \\
 B' = & \frac{2}{q_5} (P_1^2 + P_2^2) + \frac{2}{K} (P_2 P_1 + P_2 P_4) + \frac{2}{K q_5} \\
 & \sqrt{P_3^2 + P_4^2 + 2K\epsilon_1 P_1 + 2K\epsilon_2 P_2} C'
 \end{aligned} \tag{79}$$

The new coordinates  $Q_i$  are then obtained from

$$Q_i = \partial S / \partial P_i$$

which may be solved to yield the original coordinates in terms of the new as

$$\begin{aligned}
 q_1 &= Q_1 - g_1(Q_5) \\
 q_2 &= Q_2 - g_2(Q_5) \\
 q_3 &= Q_3 + Q_5 \left[ Q_1 - g_1(Q_5) \right] - \frac{f_1}{k}(Q_5) \\
 q_4 &= Q_4 + Q_5 \left[ Q_2 - g_2(Q_5) \right] - \frac{f_2}{k}(Q_5) \\
 q_5 &= -KQ_5
 \end{aligned} \tag{80}$$

Where

$$g_1(Q_5) = \partial G / \partial P_1$$

$$g_2(Q_5) = \partial G / \partial P_2$$

$$f_1(Q_5) = \partial G / \partial P_3$$

$$f_2(Q_5) = \partial G / \partial P_4$$

$$\text{where } G = \frac{F}{K} \left\{ C'(Q_5) + \frac{P_1 P_3 + P_2 P_4}{\sqrt{P_3^2 + P_4^2 + 2K\epsilon_1 P_1 + 2K\epsilon_2 P_2}} \ln A'(Q_5) - \sqrt{P_1^2 + P_2^2} \ln B' \right\}$$

Finally, the guidance function is obtained as

$$\tan \chi = \frac{P_1 - P_3 Q_5 + \epsilon_1 K Q_5^2}{P_2 - P_4 Q_5 + \epsilon_2 K Q_5^2} \quad (81)$$

A bi-quadratic form which becomes in terms of time by replacing  $Q_5$

by its solution

$$Q_5 = t + \tau = t - \frac{M_0}{K}$$

Then

$$\tan \chi = \frac{\epsilon_1 K t^2 - (2M_0 \epsilon_1 + P_3)t + (\epsilon_1 + P_3) \frac{M_0}{K} + P_1}{\epsilon_2 K t^2 - (2M_0 \epsilon_2 + P_4)t + (\epsilon_2 + P_4) \frac{M_0}{K} + P_2} \quad (81a)$$

which is an expression containing two unknown constants which may

be used to "fit" known solutions for guidance purposes.

## CONCLUSION

The application of the Hamilton-Jacobi theory of Classical Mechanics was useful in obtaining solutions to both the zero gravity and the flat earth-constant gravity rocket flight problems. These solutions then led to a first order approximate solution of the inverse square gravitational attraction problem. However, the theory did not prove useful in obtaining a closed form solution to the inverse square problem.

The development of a closed form solution by these methods depends on the proper choice of coordinates to insure that the Hamilton-Jacobi equation is separable or solvable. Consequently, it appears that the usefulness of these methods in high thrust applications will be limited until the development of a transformation procedure which will transform the system from the well known cartesian or polar coordinates to a system of coordinates for which a solution of the Hamilton-Jacobi equation is guaranteed.



## REFERENCES

1. Pontryagin, L. S., V. G. Boltyanskii, R. V. Gamkrelidze and E. F. Mishchenko; "The Mathematical Theory of Optimal Processes," John Wiley and Sons, New York, 1962.
2. Goldstein, H., "Classical Mechanics," Addison-Wesley Publishing Co. Inc., Reading, Mass., 1950.
3. Courant, R. and D. Hilbert, "Methods of Mathematical Physics, Volume II, Partial Differential Equations," John Wiley and Sons, New York, 1962.
4. Smart, V. M., "Celestial Mechanics," John Wiley and Sons, New York, 1953.
5. Miller, Frederic H., "Partial Differential Equations," John Wiley and Sons, New York, 1941.
6. Frank, P. and von Mises, "Die Differential- und Integralgleichungen der Mechanik und Physik."

HAYES INTERNATIONAL CORPORATION

A FIRST ORDER DELAUNAY SOLUTION  
FOR MINIMUM FUEL,  
LOW THRUST TRAJECTORIES

BY  
HARRY PASSMORE, III

BIRMINGHAM, ALABAMA

N65 33060 297

## SUMMARY

33060

This paper utilizes the similarity of the minimum fuel trajectory equations to those representing a restricted three-body problem to gain a canonical formulation in the variables of Delaunay. A two step transform procedure carried to the first order in small parameters is then presented as an indication of a method that may be followed in higher order studies. This progress report presents the analytical development of the procedure as completed through December, 1964.

*Author*

## INTRODUCTION

The minimum fuel equations of motion are synthesized from a generalized Hamiltonian using Pontryagin's method. These equations are, of course, identical to those presented in Reference 1 which are developed through calculus of variations procedures. Examination of the multiplier equations reveals that they may conceptually be considered as representing the motion of another (fictitious) body relative to the vehicle. A transformation of the coordinates then yields equations relative to a common center with the vehicle position coordinates.

These equations are then in a form quite similar to equations representing a three body problem in cartesian coordinates. Hence, they are easily transformed into perturbation equations in elliptic coordinates and thereby into canonical equations in a set of variables representative of those used by Delaunay in his lunar studies.

The disturbing functions of both sets of equations are not identical. However, the disturbing function of one set may be separated into two parts, one part of which is identical to the disturbing function of the other set. Two basic transforms may then be performed which shift the periodic terms into terms whose coefficients contain higher orders of small parameters. The method used by Delaunay is not applied directly. Instead, a procedure, similar to that attributed by Poincare to Bohlin, which makes use of a determining function to obtain the solution to the desired order is utilized.

The complexity of the problem and the magnitude of the task of expanding the forcing functions and obtaining the transformed relations precludes a blind approach to a higher order solution. Hence, a first order solution, as presented here, will be employed in an effort to gain insight into the order of solution required to achieve the accuracy required in space flight trajectory calculations.

## THE EQUATIONS

The two dimensional equations of motion of a constant thrust vehicle in an inverse square gravitational field may be expressed in first order form as

$$\begin{aligned}
 \dot{x}_4 &= -\frac{\mu}{r^3} x_1 + \frac{T}{m} \cos \chi \\
 \dot{x}_5 &= -\frac{\mu}{r^3} x_2 + \frac{T}{m} \sin \chi \\
 \dot{x}_1 &= x_4 \\
 \dot{x}_2 &= x_5 \\
 \dot{m} &= -\xi
 \end{aligned} \tag{1}$$

A generalized Hamiltonian function may be formulated from equations (1) as

$$\begin{aligned}
 H = \lambda_1 x_4 + \lambda_2 x_5 - \frac{\mu}{r^3} (\lambda_4 x_1 + \lambda_5 x_2) + \frac{T}{m} (\lambda_4 \cos \chi + \lambda_5 \sin \chi) \\
 - \lambda_7 \xi \quad \text{where } r^2 = x_1^2 + x_2^2
 \end{aligned}$$

to obtain the optimum thrust direction requires that

$$\begin{aligned}
 \frac{\partial H}{\partial \chi} &= 0 = -\lambda_4 \sin \chi + \lambda_5 \cos \chi \\
 \tan \chi &= \lambda_5 / \lambda_4
 \end{aligned}$$

From which

$$\sin \chi = \frac{\lambda_5}{\rho}; \quad \cos \chi = \frac{\lambda_4}{\rho}$$

where  $\rho = (\lambda_4^2 + \lambda_5^2)^{1/2}$

Substituting these values for  $\sin \chi$  and  $\cos \chi$ , H becomes

$$H = \lambda_1 x_4 + \lambda_2 x_5 - \frac{\mu}{r^3} (\lambda_4 x_1 + \lambda_5 x_2) + \frac{T}{m} \rho + \lambda_7 \alpha$$

$$\text{where } \frac{T}{m} = \frac{T}{M_0 - \xi \cdot t} = \frac{T/M_0}{1 - \frac{\xi}{M_0} t}$$

$$\text{and } x_7 = 1 - \frac{\xi}{M_0} t = 1 + \alpha t$$

$$T/M_0 = f$$

$$\frac{T}{m} = f/x_7 \quad \bar{\lambda}_7 = M_0 \lambda_7$$

The equations which must be solved to obtain minimum fuel trajectories become

$$\begin{aligned} \dot{x}_4 &= \partial H / \partial \lambda_4 = -\frac{\mu}{r^3} x_1 + \frac{f}{x_7} \frac{\lambda_4}{\rho} \\ \dot{x}_5 &= \partial H / \partial \lambda_5 = -\frac{\mu}{r^3} x_2 + \frac{f}{x_7} (\lambda_5 / \rho) \\ \dot{x}_1 &= \partial H / \partial \lambda_1 = x_4 \\ \dot{x}_2 &= \partial H / \partial \lambda_2 = x_5 \\ \dot{x}_7 &= \partial H / \partial \lambda_7 = \alpha \\ \dot{\lambda}_4 &= -\partial H / \partial x_4 = -\lambda_1 \\ \dot{\lambda}_5 &= -\partial H / \partial x_5 = -\lambda_2 \\ \dot{\lambda}_1 &= -\partial H / \partial x_1 = \frac{\mu}{r^3} \lambda_4 - \frac{3\mu x_1}{r^5} (\lambda_4 x_1 + \lambda_5 x_2) \\ \dot{\lambda}_2 &= -\partial H / \partial x_2 = \frac{\mu}{r^3} \lambda_5 - \frac{3\mu x_2}{r^5} (\lambda_4 x_1 + \lambda_5 x_2) \\ \dot{\lambda}_7 &= -\partial H / \partial x_7 = \frac{f}{x_7^2} \rho \end{aligned} \tag{2}$$

Now, returning to the expressions for  $\sin \chi$  and  $\cos \chi$ , and referring to Figure 1, it may be seen that the thrust direction may be considered as the direction to some fictitious body a distance  $\rho$  from the vehicle.  $\lambda_4$  and  $\lambda_5$  may then be considered as the coordinates parallel to  $x_1$  and  $x_2$  of the fictitious body relative to the vehicle. To obtain equations analogous to three body equations, the  $\lambda$  equations must be transformed to equations relative to the same center of attraction as the vehicle. This may be accomplished by introducing

$$\begin{aligned}
\psi_4 &= x_1 + \lambda_4 \\
\psi_5 &= x_2 + \lambda_5 \\
\psi_7 &= \bar{\lambda}_7
\end{aligned} \tag{3}$$

such that  $\rho$  is now

$$\rho = \left[ (\psi_4 - x_1)^2 + (\psi_5 - x_2)^2 \right]^{1/2}$$

or upon defining

$$\Delta^2 = \psi_4^2 + \psi_5^2$$

$\rho$  becomes

$$\rho = \left[ \Delta^2 - 2(x_1 \psi_4 + x_2 \psi_5) + r^2 \right]^{1/2}$$

The equations of motion (2) are then transformed to the following second order equations

$$\begin{aligned}
\ddot{x}_1 + \frac{\mu x_1}{r^3} &= \frac{f}{x_7 \rho} (\psi_4 - x_1) \\
\ddot{x}_2 + \frac{\mu x_2}{r^3} &= \frac{f}{x_7 \rho} (\psi_5 - x_2) \\
\ddot{\psi}_4 &= \frac{f}{x_7 \rho} (\psi_4 - x_1) - \frac{\mu \psi_4}{r^3} + \frac{3\mu x_1}{r^5} [x_1 \psi_4 + x_2 \psi_5 - r^2] \\
\ddot{\psi}_5 &= \frac{f}{x_7 \rho} (\psi_5 - x_2) - \frac{\mu \psi_5}{r^3} + \frac{3\mu x_2}{r^5} [x_1 \psi_4 + x_2 \psi_5 - r^2]
\end{aligned}$$

Examining the right hand sides of these equations and defining

$$\begin{aligned}
R_1 &= -\frac{f}{x_7} \rho \\
R_2 &= \frac{f}{x_7} \rho - \frac{\mu}{\Delta} - \mu \frac{\Delta^2}{2r^3} - \frac{3\mu}{r^5} (x_1 \psi_4 + x_2 \psi_5) \left[ 1 - \frac{x_1 \psi_4 + x_2 \psi_5}{2r^2} \right]
\end{aligned}$$

The equations of interest become

$$\begin{aligned}
\ddot{x}_1 + \mu \frac{x_1}{r^3} &= \partial R_1 / \partial x_1 & \ddot{\psi}_4 + \frac{\mu \psi_4}{\Delta^3} &= \partial R_2 / \partial \psi_4 \\
\ddot{x}_2 + \mu \frac{x_2}{r^3} &= \partial R_1 / \partial x_2 & \ddot{\psi}_5 + \frac{\mu \psi_5}{\Delta^3} &= \partial R_2 / \partial \psi_5
\end{aligned} \tag{4}$$

These equations are then identical in form to equations representing a restricted three body problem and may be transformed by any of several standard methods available into canonical equations in the Delaunay variables.



$$\begin{aligned}
\dot{L}_v &= \partial F_v / \partial \ell_v & \dot{\ell}_v &= - \partial F_v / \partial L_v \\
\dot{G}_v &= \partial F_v / \partial g_v & \dot{g}_v &= - \partial F_v / \partial G_v \\
\dot{K} &= \partial F_v / \partial k & \dot{k} &= - \partial F_v / \partial K \\
\dot{L}_T &= \partial F_T / \partial \ell_T & \dot{\ell}_T &= - \partial F_T / \partial L_T \\
\dot{G}_T &= \partial F_T / \partial g_T & \dot{g}_T &= - \partial F_T / \partial G_T
\end{aligned}$$

where

$$\begin{aligned}
F_v &= \frac{\mu^2}{2L_v^2} - \alpha K + R_1 \\
F_T &= \frac{\mu^2}{2L_T^2} + \alpha K + R_2
\end{aligned}$$

and the substitution  $x_7 = k$ ,  $\psi_7 = K$  has been incorporated to account for the mass equation. The subscript v applies to parameters which represent the vehicle and the subscript T indicates parameters representing the thruster body. Now, upon adding  $-\frac{\mu^2}{2L_T^2}$  to  $F_v$  and  $-\frac{\mu^2}{2L_v^2}$  to  $F_T$ , substituting the values for  $R_1$  and  $R_2$  and defining

$$F_2 = -\mu \left\{ \frac{1}{\Delta} + \frac{\Delta^2}{2r^3} + \frac{3}{r^3} (x_1 \psi_4 + x_2 \psi_5) - \frac{3}{2r^5} (x_1 \psi_4 + x_2 \psi_5)^2 \right\}$$

$F_v$  and  $F_T$  may be expressed

$$\begin{aligned}
F_v &= \frac{\mu^2}{2L_v^2} - \frac{\mu^2}{2L_T^2} - \alpha K - \frac{f}{k} \rho \\
F_T &= -F_v + F_2
\end{aligned}$$

and the equations may be expressed

$$\begin{aligned}
\dot{L}_v &= \partial F_v / \partial \ell_v & \dot{\ell}_v &= - \partial F_u / \partial L_v \\
\dot{G}_v &= \partial F_v / \partial g_v & \dot{g}_v &= - \partial F_v / \partial G_v \\
\dot{K} &= \partial F_v / \partial k & \dot{k} &= - \partial F_v / \partial K \\
\dot{L}_T &= - \partial F_v / \partial \ell_T + \partial F_2 / \partial \ell_T; \dot{\ell}_T &= + \partial F_v / \partial L_T - \partial F_2 / \partial L_T \\
\dot{G}_T &= - \partial F_v / \partial g_T + \partial F_2 / \partial g_T; \dot{g}_T &= + \partial F_v / \partial L_T - \partial F_2 / \partial L_T
\end{aligned} \tag{5}$$

## THE DISTURBING FUNCTION EXPANSIONS

To obtain solutions of equations (5) by applying a Delaunay procedure, it is necessary to express  $F_v$  and  $F_2$  as series expansions in the Delaunay variables  $L$ ,  $G$ ,  $\ell$  and  $g$  and/or the closely related elliptic parameters,  $a$  and  $e$ . These types of expansions are readily available in any of several texts on Celestial Mechanics. The actual functions  $F_v$  and  $F_2$  of interest here are not identical with those found in the texts, however, the individual parameters in the functions are similar and the expansion procedures are the same. Therefore, only the results of the expansions taken to the first order in the eccentricity  $e$  will be presented here.

The expanded form for  $F_v$  is then

$$\begin{aligned}
 F_v = & \frac{\mu^2}{2L_v^2} - \frac{\mu^2}{2L_T^2} - \alpha K - \epsilon_1 \left\{ .85 (L_v^4 + L_T^4)^{1/2} \right. \\
 & + .065 L_v L_T (L_v^4 + L_T^4)^{-1/2} (L_v + G_v)(L_T + G_T) \left[ \cos(\ell_v + g_v + k - \ell_T - g_T) \right. \\
 & \quad \left. + \cos(\ell_v + g_v - k - \ell_T - g_T) \right] \\
 & + .032 L_v L_T (L_v^4 + L_T^4)^{-1/2} (L_v + G_v)(L_T + G_T) \left[ \cos(\ell_v + g_v + 2k - \ell_T - g_T) \right. \\
 & \quad \left. + \cos(\ell_v + g_v - 2k - \ell_T - g_T) \right] \\
 & + .637 e_v L_v^2 L_T (L_v^4 + L_T^4)^{-1/2} (L_T + G_T) \cos(-g_v + \ell_T + g_T) \\
 & + .08 e_T L_v L_T^2 (L_v + G_v)(L_v^4 + G_v^4)^{-3/2} \left[ 8L_v^4 + 7L_T^4 + L_T^3 G_T \right] \cos(\ell_v + g_v - g_T) \\
 & - .258 e_T L_T^4 (L_v^4 + L_T^4)^{-1/2} \left[ \cos(k + \ell_T) + \cos(k - \ell_T) \right] \\
 & + .194 e_v L_v^2 L_T (L_T + G_T)(L_v^4 + L_T^4)^{-1/2} \left[ \cos(-g_v + k + \ell_T + g_T) + \cos(-g_v - k + \ell_T + g_T) \right] \\
 & + e_T L_v L_T^2 (L_v + G_v)(L_T^4 + L_v^4)^{-3/2} \left[ .194 L_v^4 + .19 L_T^4 - .04 L_T^3 G_T \right] \\
 & \quad \times \left[ \cos(\ell_v + g_v + k - g_T) + \cos(\ell_v + g_v - k - g_T) \right] \\
 & - .85 e_T L_T^4 (L_v^4 + L_T^4)^{-1/2} \cos \ell_T
 \end{aligned}$$

$$\begin{aligned}
& -.214 L_v L_T (L_v^4 + L_T^4)^{-1/2} (L_v + G_v)(L_T + G_T) \cos (\ell_v + g_v - \ell_T - g_T) \\
& + .52 (L_v^4 + L_T^4)^{1/2} \cos k \\
& + .127 (L_v^4 + L_T^4)^{1/2} \cos 2k \\
& + .057 (L_v^4 + L_T^4)^{1/2} \cos 3k \\
& + .063 (L_v^4 + L_T^4)^{1/2} \cos 4k \}
\end{aligned}$$

where  $\epsilon_1 = f/\mu$

Likewise, the expansion, to the first order in the eccentricities, of

$F_2$  is

$$\begin{aligned}
F_2 = & -\frac{\eta^2}{\mu^2} \left\{ L_T^{-2} \left[ \frac{1}{2} L_T^6 + L_v^6 - \frac{3}{64} (L_v + G_v)^2 L_v^{-2} L_T^4 (L_T + G_T)^2 \right] \right. \\
& + e_v \left[ \frac{3}{2} L_T^2 + \frac{3}{32} L_v^{-1} L_T^2 (L_T + G_T)^2 (L_v + G_v) (3 - L_v^{-1} (L_v + G_v)) \right] \cos \ell_v \\
& + e_T L_T^{-2} \left[ L_v^6 + L_T^6 + \frac{3}{64} L_v^{-2} L_T^4 (L_v + G_v)^2 (L_T + G_T) (5L_T - G_T) \right] \cos \ell_T \\
& - \frac{3}{64} L_v^{-2} L_T^2 (L_v + G_v)^2 (L_v + G_T)^2 \cos (2\ell_v + 2g_v - 2\ell_T - 2g_T) \\
& + \frac{3}{128} e_v L_v^{-1} L_T^2 (L_v + G_v) (L_T^2 + G_T^2) \left[ 12 - L_v^{-1} (L_v + G_v) \right] \cos (\ell_v + 2g_v - 2\ell_T - 2g_T) \\
& + \frac{9}{32} e_T L_v^{-2} L_T^3 (L_v + G_v)^2 (L_T + G_T) \cos (2\ell_v + 2g_v - \ell_T - 2g_T) \\
& - \frac{3}{64} e_T L_v^{-2} L_T^2 (L_v + G_v)^2 (L_T + G_T)^2 \cos (2\ell_v + 2g_v - 3\ell_T - 2g_T) \\
& \left. - \frac{9}{128} e_v L_v^{-2} L_T^2 (L_v + G_v)^2 (L_T + G_T)^2 \cos (3\ell_v + 2g_v - 2\ell_T - 2g_T) \right\}
\end{aligned}$$

where  $\eta = \frac{\mu^2}{L_v^3}$

## THE FIRST TRANSFORM

A previous section presented the canonical equations and the expressions for the forcing functions in terms of the variables  $L_v$ ,  $G_v, l_v, g_v, K, k$  for the vehicle and  $L_T, G_T, l_T, g_T$  for the thruster. To aid in the bookkeeping in the transformations and to achieve a slight realignment of the equations, the following notation is introduced.

$$\begin{array}{ll}
 L_v = L_{10} & l_v = l_{10} \\
 G_v = L_{20} & g_v = l_{20} \\
 K = L_{30} & k = l_{30} \\
 L_T = -L_{40} & l_T = l_{40} \\
 G_T = -L_{50} & g_T = l_{50}
 \end{array}$$

also, let  $F_v = F_1$  such that

$$F_T = -F_1 + F_2$$

The equations of interest, equations 5, then become

$$\begin{array}{ll}
 \dot{L}_{10} = \partial F_1 / \partial l_{10} & \dot{l}_{10} = -\partial F_1 / \partial L_{10} \\
 \dot{L}_{20} = \partial F_1 / \partial l_{20} & \dot{l}_{20} = -\partial F_1 / \partial L_{20} \\
 \dot{L}_{30} = \partial F_1 / \partial l_{30} & \dot{l}_{30} = -\partial F_1 / \partial L_{30} \\
 \dot{L}_{40} = \partial F_1 / \partial l_{40} - \partial F_2 / \partial l_{40} & \dot{l}_{40} = -\partial F_1 / \partial L_{40} + \partial F_2 / \partial L_{40} \\
 \dot{L}_{50} = \partial F_1 / \partial l_{50} - \partial F_2 / \partial l_{50} & \dot{l}_{50} = -\partial F_1 / \partial L_{50} + \partial F_2 / \partial L_{50}
 \end{array} \quad (5a)$$

The equations for the first transform are obtained by neglecting the term  $F_2$  in the above expressions.  $\dot{L}_{j0} = \partial F_1 / \partial l_{j0}$ ;  $\dot{l}_{j0} = -\frac{\partial F_1}{\partial L_{j0}}$  (6)

Then expressing

$$F_1 = F_{10} + F_{11} \quad (j = 1, \dots, 5)$$

where  $F_{10}$  is the part of  $F_1$  that does not contain the small parameter  $\epsilon_1$  and may be seen from the function expansions to be

$$F_{10} = \frac{\mu^2}{2L_{10}^2} - \frac{\mu^2}{2L_{40}^2} - \alpha L_{30} \quad (7)$$

The term  $F_{11}$  consists of all terms in  $F_1$  which contain the small parameter  $\epsilon_1$  and may be expressed as

$$F_{11} = P_0 + \sum_{i=1}^n Q_{0i} \cos \theta_{0i} \quad (8)$$

where  $P_0$  is the part of  $F_{11}$  that contains no periodic terms and may be seen to be

$$P_0 = - (.85) \epsilon_1 (L_{10}^4 + L_{40}^4)^{1/2}$$

$Q_{0i}$  represents the coefficients of the periodic terms and as may again be seen from the expression for  $F_v$  in the expansion section, the  $Q_{0i}$  are functions of the small parameter  $\epsilon_1$  and  $L_{10}$ ,  $L_{20}$ ,  $L_{40}$  and  $L_{50}$  only.

The  $\cos \theta_{0i}$  are the periodic terms where the  $\theta_{0i}$  are given by

$$\theta_{0i} = p_{1i} l_{10} + p_{2i} l_{20} + p_{3i} l_{30} + p_{4i} l_{40} + p_{5i} l_{50} \quad (9)$$

or

$$\theta_{0i} = \sum_{j=1}^5 p_{ji} l_{j0} \quad (9a)$$

and  $n$  is the number of periodic terms to be considered.

The procedure now is to transform the Hamiltonian of this part of the problem,  $F_1$ , into a new Hamiltonian which is independent of the angle variables such that

$$\begin{aligned} F_1 (L_{10}, L_{20}, L_{30}, L_{40}, L_{50}, l_{10}, l_{20}, l_{30}, l_{40}, l_{50}) \\ = F_1^* (L_{11}, L_{21}, L_{31}, L_{41}, L_{51}) \end{aligned} \quad (10)$$

where  $L_{11}$ ,  $L_{21}$ , etc., represent the transformed variables.

To aid in the transformation, a determining function

$$S = S(L_{j1}, l_{j0}) \quad (j = 1, 2, \dots, 5)$$

may be used following the procedures of the Hamilton-Jacobi theory.

The equations of transformation are then

$$\begin{aligned} L_{j0} &= \partial S / \partial l_{j0} \\ l_{j1} &= \partial S / \partial L_{j1} \end{aligned} \quad (j = 1, 2, \dots, 5) \quad (11)$$

and the terms of the Hamiltonian become

$$\begin{aligned} F_{10} &= F_{10} \left( \partial S / \partial l_{10}, \frac{\partial S}{\partial l_{30}}, \frac{\partial S}{\partial l_{40}} \right) \\ F_{11} &= F_{11} \left( \frac{\partial S}{\partial l_{10}}, \frac{\partial S}{\partial l_{20}}, \frac{\partial S}{\partial l_{40}}, \frac{\partial S}{\partial l_{50}}, l_{10}, l_{20}, l_{30}, l_{40}, l_{50} \right) \end{aligned}$$

The determining function may be expanded in powers of the small parameter  $\epsilon_1$ , as

$$S = S_0 + S_1 + S_2 + \dots$$

where  $S_0$  does not contain  $\epsilon_1$ ,  $S_1$  is first order in  $\epsilon_1$ ,  $S_2$  is second order etc.

To insure an identity transformation in case all  $Q_{01}$  happen to be zero,  $S_0$  must be

$$S_0 = L_{11}l_{10} + L_{21}l_{20} + L_{31}l_{30} + L_{41}l_{40} + L_{51}l_{50} \quad (12)$$

or 
$$S_0 = \sum_j L_{j1}l_{j0}$$

The transformed Hamiltonian may also be expanded in powers of the small parameter  $\epsilon_1$  as

$$F_1^* = F_{10}^* + F_{11}^* + F_{12}^* + \dots \quad (13)$$

where  $F_{10}^*$  is of zero order in  $\epsilon_1$ ,  $F_{11}^*$  is of first order, etc. Substituting the relations for  $S$  and  $F_1^*$ , the Hamilton-Jacobi equation becomes

upon substitution of equations 11, 12, and 13 in 10

$$F_0 \left[ L_{j1} + \left( \frac{\partial S_1}{\partial \ell_{j0}} + \frac{\partial S_2}{\partial \ell_{j0}} + \dots \right) \right] + F_1 \left[ L_{j1} + \left( \frac{\partial S_1}{\partial \ell_{j0}} + \frac{\partial S_2}{\partial \ell_{j0}} + \dots \right) \right. \\ \left. \theta_{01} \right] = F_{10}^* + F_{11}^* + \dots \quad (14)$$

$$(j = 1, 2, \dots, 5)$$

$F_0$  and  $F_1$  may then be expanded in Taylor's series which to the first order in  $\epsilon_1$  become

$$F_0 = F_0(L_{j1}) + \sum_j \left. \frac{\partial F_0}{\partial L_{j0}} \right|_{L_{j1}} \left( \frac{\partial S_1}{\partial \ell_{j0}} + \dots \right)$$

$$F_1 = F_1(L_{j1}, \theta_{01}) \quad (j = 1, 2, \dots, 5)$$

$$(i = 1, 2, \dots, n)$$

Substituting these series into the Hamilton-Jacobi equation and equating terms of like order in  $\epsilon_1$  gives

$$F_{10}(L_{j1}) = F_{10}^* = \frac{\mu^2}{2L_{11}^2} - \frac{\mu^2}{2L_{41}^2} - \alpha L_{31} \quad (15a)$$

$$\sum_j \left. \frac{\partial F_{10}}{\partial L_{j0}} \right|_{L_{j1}} \frac{\partial S_1}{\partial \ell_{j0}} + P_1 + \sum_i Q_{1i} \cos \theta_{0i} = F_{11}^*(L_{j1}) \quad (15b)$$

The notation  $\left. \frac{\partial F_{10}}{\partial L_{j0}} \right|_{L_{j1}}$  denotes  $\frac{\partial F_{10}}{\partial L_{j0}}$  evaluated at  $L_{j0} = L_{j1}$ ,  $P_1$  and  $Q_{1i}$  denote the functions  $P_0$  and  $Q_{0i}$  with the  $L_{j1}$  substituted for the  $L_{j0}$ . Now, since  $F_{11}^*$  is a function of the  $L_{j1}$  only and since  $S_1$  and  $\sum_i Q_{1i} \cos \theta_{0i}$  are functions of the  $\ell_j$ ,  $F_{11}^*$  can only be related to the term  $P_1$ . Hence,

$$F_{11}^*(L_{j1}) = P_1 = - (.85) \epsilon_1 (L_{11}^4 + L_{41}^4)^{1/2} \quad (16)$$

and

$$\sum_j \left. \frac{\partial F_{10}}{\partial L_{j0}} \right|_{L_{j1}} \frac{\partial S_1}{\partial \ell_{j0}} = - \sum_i Q_{1i} \cos \theta_{0i} \quad (17)$$

Returning to the expression for  $F_{10}$ , equation (7), and introducing the notation

$$v_{10} = \frac{\mu^2}{L_{10}^3}, \quad v_{40} = \frac{\mu^2}{L_T^3} = -\frac{\mu^2}{L_{40}^3} \quad (18)$$

The required derivatives may be evaluated as

$$\left. \frac{\partial F_{10}}{\partial L_{10}} \right|_{L_{11}} = - \left. \frac{\mu^2}{L_{10}^3} \right|_{L_{11}} = - v_{10} \Big|_{L_{11}} = -v_{11}$$

$$\left. \frac{\partial F_{10}}{\partial L_{20}} \right|_{L_{21}} = 0$$

$$\left. \frac{\partial F_{10}}{\partial L_{30}} \right|_{L_{31}} = -\alpha = -v_{31}$$

$$\left. \frac{\partial F_{10}}{\partial L_{40}} \right|_{L_{41}} = \left. \frac{\mu^2}{L_{40}^3} \right|_{L_{41}} = -v_{40} \Big|_{L_{41}} = -v_{41}$$

$$\left. \frac{\partial F_{10}}{\partial L_{50}} \right|_{L_{51}} = 0$$

or in general

$$\left. \frac{\partial F_{10}}{\partial L_{j0}} \right|_{L_{j1}} = -v_{j1}$$

The equation for  $S_1$ , equation (17), then becomes

$$v_{11} \frac{\partial S_1}{\partial l_{10}} + \alpha \frac{\partial S_1}{\partial l_{30}} + v_{41} \frac{\partial S_1}{\partial l_{40}} = \sum_{i=1}^n Q_{1i} \cos \theta_{0i} \quad (19)$$

or

$$\sum_j v_{j1} \partial S_1 / \partial L_{j0} = \sum_i Q_{1i} \cos \theta_{0i} \quad (19a)$$

A solution for this equation may be taken in the form

$$S_1 = \sum_i A_{1i} \sin \theta_{0i} \quad (20)$$

where the  $A_{1i}$  are not functions of any  $l_{j0}$ . Hence,

$$\frac{\partial S_1}{\partial l_{j0}} = \sum_i \frac{\partial \theta_{0i}}{\partial l_{j0}} A_{1i} \cos \theta_{0i}$$



The particles of  $\theta_{01}$  required are obtained from the expression for  $\theta_{01}$ , equation (9), as

$$\frac{\partial \theta_{01}}{\partial l_{10}} = p_{1i} ; \quad \frac{\partial \theta_{01}}{\partial l_{30}} = p_{3i} ; \quad \frac{\partial \theta_{01}}{\partial l_{40}} = p_{4i}$$

or in general

$$\frac{\partial \theta_{01}}{\partial l_{j0}} = p_{ji}$$

Substitution of the assumed solution into the equation then gives

$$\sum_i A_{1i} \sum_j v_{ji} p_{ji} \cos \theta_{01} = \sum_i Q_{1i} \cos \theta_{01}$$

Equating coefficients of like cosines yields

$$A_{1i} = \frac{Q_{1i}}{\sum_j v_{ji} p_{ji}} \quad (i = 1, \dots, n) \quad (21)$$

Hence

$$S_1 = \sum_{i=1}^n \frac{Q_{1i}}{\sum_j v_{ji} p_{ji}} \sin \theta_{01} \quad (22)$$

With these values for  $S_0$  and  $S_1$ , the determining function to the first order in  $\epsilon_1$  becomes

$$S = \sum_j L_{j1} l_{j0} + \sum_i \frac{Q_{1i}}{\sum_j v_{ji} p_{ji}} \sin \theta_{01} \quad (23)$$

The equations of transform then give

$$L_{j0} = \frac{\partial S}{\partial l_{j0}} = L_{j1} + \sum_i A_{1i} p_{ji} \cos \theta_{01} \quad (j = 1, \dots, 5) \quad (24)$$

$$l_{j1} = \partial S / \partial L_{j1} = l_{j0} + \sum_i \frac{\partial A_{1i}}{\partial L_{j1}} \sin \theta_{01}$$

The complete first order Hamiltonian function for this part of the problem may now be written in the new variables as

$$F_1^* = \frac{\mu^2}{2L_{11}^2} - \frac{\mu^2}{2L_{41}^2} - \alpha L_{31} - .85\epsilon_1 (L_{11}^4 + L_{41}^4)^{1/2} \quad (28)$$

The corresponding equations of motion are then

$$\dot{L}_{j1} = \partial F_1^* / \partial l_{j1} ; \dot{l}_{j1} = - \partial F_1^* / \partial L_{j1} \quad (j = 1, \dots, 5) \quad (29)$$

The solutions of (29) may then be written

$$\begin{aligned} L_{j1} &= a_j \\ l_{j1} &= f_j t + b_j \end{aligned} \quad (j = 1, \dots, 5) \quad (30)$$

where  $a_j$ ,  $f_j$  and  $b_j$  are constants.

Substituting the values of  $\ell_{j0}$  from these equations into the expression for  $\theta_{0i}$  equation ( 9a ), gives

$$\theta_{0i} = \sum_{j=1}^5 p_{ji} \left[ \ell_{ji} - \sum_{k=1}^n \frac{\partial A_{1k}}{\partial L_{j1}} \sin \theta_{0k} \right]$$

Upon defining

$$\theta_{1i} = \sum_{j=1}^5 p_{ji} \ell_{ji} \quad (25)$$

$\theta_{0i}$  becomes

$$\theta_{0i} = \theta_{1i} - \sum_{j=1}^5 p_{ji} \sum_{k=1}^n \frac{\partial A_{1k}}{\partial L_{j1}} \sin \theta_{0k} \quad (26)$$

where the index of the second summation has been changed to avoid confusion. This expression may then be written

$$\begin{aligned} \theta_{0i} = \theta_{1i} - \sum_{j=1}^5 p_{ji} & \left[ \sum_{k=1}^{i-1} \frac{\partial A_{1k}}{\partial L_{j1}} \sin \theta_{0k} + \sum_{k=i+1}^n \frac{\partial A_{1k}}{\partial L_{j1}} \sin \theta_{0k} \right] \\ & - \sum_{j=1}^5 p_{ji} \frac{\partial A_{1i}}{\partial L_{j1}} \sin \theta_{0i} \end{aligned}$$

which is in a form to which the Lagrange expansion theorem is applicable. Applying this theorem and performing the necessary simplifications gives the values for  $\cos \theta_{0i}$  and  $\sin \theta_{0i}$  needed in the transform equations.

$$\cos \theta_{0i} = \cos \theta_{1i} + (\text{terms of first and higher order in } \epsilon_1)$$

$$\sin \theta_{0i} = \sin \theta_{1i} + (\text{terms of first and higher order in } \epsilon_1)$$

Then, since  $A_{1i}$  is itself a quantity of first order in  $\epsilon_1$ , the transformation equations become

$$\begin{aligned} L_{j0} &= L_{j1} + \sum_i A_{1i} p_{ji} \cos \theta_{1i} \quad j = 1, 2, \dots, 5 \quad (27) \\ \ell_{j0} &= \ell_{j1} - \sum_i E_{ji} \sin \theta_{1i} \quad \text{where } E_{ji} = \partial A_{1i} / \partial L_{j1} \end{aligned}$$

These expressions are of course a great deal more complicated when higher order solutions are sought.

## DEVELOPMENT OF AN APPROXIMATE CANONICAL FORM

Substitution of the solution equations 30 or the transformation equations 27 will not yield equations in a canonical form in  $F_2$ , directly. Therefore, it is necessary to make further small order approximations to obtain equations in a form suitable for further application of the procedure. To illustrate this, and to provide a somewhat simplified outline of the developments performed in the first transformation, consider the equations of motion 5a in the following form:

$$\begin{aligned} \dot{L}_{po} &= \partial F_1 / \partial l_{po} & \dot{l}_{po} &= - \partial F_1 / \partial L_{po} \\ \dot{L}_{qo} &= \frac{\partial F_1}{\partial l_{qo}} - \frac{\partial F_2}{\partial l_{qo}} & \dot{l}_{qo} &= - \partial F_1 / \partial L_{qo} + \partial F_2 / \partial L_{qo} \end{aligned} \quad (31)$$

$$(p = 1, 2, 3) \quad (q = 4, 5)$$

$$F_1 = F_1 (L_{po}, l_{po}, L_{qo}, l_{qo}) \quad F_2 = F_2 (L_{po}, l_{po}, L_{qo}, l_{qo})$$

The technique followed so far has been to obtain solutions to the equations obtained by neglecting  $F_2$ .

$$\begin{aligned} \dot{L}_{jo} &= \partial F_1 / \partial l_{jo} & \dot{l}_{jo} &= \frac{-\partial F_1}{\partial L_{jo}} \end{aligned} \quad (32)$$

$$(j = 1 \text{---} 5)$$

The solution to these equations were found by solving the Hamilton-Jacobi equation

$$F_1 (l_{jo}, \partial S / \partial l_{jo}) = F_1^* (L_{ji}) \quad (33)$$

where  $S$  is the determining function

$$S = S(L_{j1}, \ell_{jo})$$

The equations of transformation were then obtained from

$$L_{jo} = \partial S / \partial \ell_{jo} \quad \ell_{j1} = \partial S / \partial L_{j1} \quad (34)$$

which gave

$$\begin{aligned} L_{jo} &= L_{jo}(L_{k1}, \ell_{k1}) = L_{j1} + \sum_i A_{ji} p_{ji} \cos \theta_{oi} \\ \ell_{jo} &= \ell_{jo}(L_{k1}, \ell_{k1}) = \ell_{j1} - \sum_i E_{ji} \sin \theta_{oi} \end{aligned} \quad (35)$$

Taking the total time derivatives of these equations and substituting

into the equations of motion 31

$$\sum_{r=1}^5 \left[ \frac{\partial L_{po}}{\partial L_{r1}} \dot{L}_{r1} + \frac{\partial L_{po}}{\partial \ell_{r1}} \dot{\ell}_{r1} \right] = \partial F_1 / \partial \ell_{po} \quad (36a)$$

$$\sum_{r=1}^5 \left[ \frac{\partial \ell_{po}}{\partial L_{r1}} \dot{L}_{r1} + \frac{\partial \ell_{po}}{\partial \ell_{r1}} \dot{\ell}_{r1} \right] = - \partial F_1 / \partial L_{po} \quad (36b)$$

$$\sum_{r=1}^5 \left[ \frac{\partial L_{qo}}{\partial L_{r1}} \dot{L}_{r1} + \frac{\partial L_{qo}}{\partial \ell_{r1}} \dot{\ell}_{r1} \right] = \partial F_1 / \partial \ell_{qo} - \partial F_2 / \partial \ell_{qo} \quad (36c)$$

$$\sum_{r=1}^5 \left[ \frac{\partial \ell_{qo}}{\partial L_{r1}} \dot{L}_{r1} + \frac{\partial \ell_{qo}}{\partial \ell_{r1}} \dot{\ell}_{r1} \right] = - \partial F_1 / \partial L_{qo} + \partial F_2 / \partial L_{qo} \quad (36d)$$

$$(p = 1, 2, 3) \quad (q = 4, 5) \quad (r = 1 - - - 5)$$

Multiplying each of equations 36a by  $\partial \ell_{po} / \partial L_{k_1}$  and each of equations 36c by  $\partial \ell_{qo} / \partial L_{k_1}$  where  $L_{k_1}$  is one particular  $L_{r_1}$  and adding gives

$$\begin{aligned}
 & \sum_{r=1}^5 \left\{ \left[ \sum_{p=1}^3 \frac{\partial L_{po}}{\partial L_{r_1}} \frac{\partial \ell_{po}}{\partial L_{k_1}} + \sum_{q=4}^5 \frac{\partial L_{qo}}{\partial L_{r_1}} \frac{\partial \ell_{qo}}{\partial L_{k_1}} \right] \dot{L}_{r_1} + \right. \\
 & \left. \left[ \sum_{p=1}^3 \frac{\partial L_{po}}{\partial \ell_{r_1}} \frac{\partial \ell_{po}}{\partial L_{k_1}} + \sum_{q=4}^5 \frac{\partial L_{qo}}{\partial \ell_{r_1}} \frac{\partial \ell_{qo}}{\partial L_{k_1}} \right] \dot{\ell}_{r_1} \right\} = \\
 & \sum_{p=1}^3 \frac{\partial F_1}{\partial \ell_{po}} \frac{\partial \ell_{po}}{\partial L_{k_1}} + \sum_q \left[ \frac{\partial F_1}{\partial \ell_{qo}} \frac{\partial \ell_{qo}}{\partial L_{k_1}} - \frac{\partial F_2}{\partial \ell_{qo}} \frac{\partial \ell_{qo}}{\partial L_{k_1}} \right] \quad (37)
 \end{aligned}$$

Multiplying each equation 36b by  $\partial L_{po} / \partial L_{k_1}$  and each of equations 36d by  $\partial L_{qo} / \partial L_{k_1}$  and adding gives

$$\begin{aligned}
 & \sum_{r=1}^5 \left\{ \left[ \sum_{p=1}^3 \frac{\partial \ell_{po}}{\partial L_{r_1}} \frac{\partial L_{po}}{\partial L_{k_1}} + \sum_{q=4}^5 \frac{\partial \ell_{qo}}{\partial L_{r_1}} \frac{\partial L_{qo}}{\partial L_{k_1}} \right] \dot{L}_{r_1} + \right. \\
 & \left. \left[ \sum_{p=1}^3 \frac{\partial L_{po}}{\partial \ell_{r_1}} \frac{\partial L_{po}}{\partial L_{k_1}} + \sum_{q=4}^5 \frac{\partial L_{qo}}{\partial \ell_{r_1}} \frac{\partial L_{qo}}{\partial L_{k_1}} \right] \dot{\ell}_{r_1} \right\} \quad (38) \\
 & = \sum_{p=1}^3 - \frac{\partial F_1}{\partial L_{po}} \frac{\partial L_{po}}{\partial L_{k_1}} - \sum_{q=4}^5 \frac{\partial F_1}{\partial L_{qo}} \frac{\partial L_{qo}}{\partial L_{k_1}} + \sum_{q=4}^5 \frac{\partial F_2}{\partial L_{qo}} \frac{\partial L_{qo}}{\partial L_{k_1}}
 \end{aligned}$$

defining brackets as

$$\begin{aligned}
 \left[ L_{k_1}, L_{r_1} \right] &= \sum_{p=1}^3 \left( \frac{\partial L_{p0}}{\partial L_{k_1}} \frac{\partial l_{p0}}{\partial L_{r_1}} - \frac{\partial L_{p0}}{\partial L_{r_1}} \frac{\partial l_{p0}}{\partial L_{k_1}} \right) \\
 &\quad + \sum_{q=4}^5 \left( \frac{\partial L_{q0}}{\partial L_{k_1}} \frac{\partial l_{q0}}{\partial L_{r_1}} - \frac{\partial L_{q0}}{\partial L_{r_1}} \frac{\partial l_{q0}}{\partial L_{k_1}} \right) \\
 \left[ L_{k_1}, l_{r_1} \right] &= \sum_{p=1}^3 \left( \frac{\partial L_{p0}}{\partial L_{k_1}} \frac{\partial l_{p0}}{\partial l_{r_1}} - \frac{\partial L_{p0}}{\partial l_{r_1}} \frac{\partial l_{p0}}{\partial L_{k_1}} \right) \\
 &\quad + \sum_{q=4}^5 \left( \frac{\partial L_{q0}}{\partial L_{k_1}} \frac{\partial l_{q0}}{\partial l_{r_1}} - \frac{\partial L_{q0}}{\partial l_{r_1}} \frac{\partial l_{q0}}{\partial L_{k_1}} \right) \\
 \left[ l_{k_1}, l_{r_1} \right] &= \sum_{p=1}^3 \left( \frac{\partial L_{p0}}{\partial l_{k_1}} \frac{\partial l_{p0}}{\partial l_{r_1}} - \frac{\partial L_{p0}}{\partial l_{r_1}} \frac{\partial l_{p0}}{\partial l_{k_1}} \right) \\
 &\quad + \sum_{q=4}^5 \left( \frac{\partial L_{q0}}{\partial l_{k_1}} \frac{\partial l_{q0}}{\partial l_{r_1}} - \frac{\partial L_{q0}}{\partial l_{r_1}} \frac{\partial l_{q0}}{\partial l_{k_1}} \right)
 \end{aligned} \tag{39}$$

and subtracting equation 37 from 38 gives

$$\begin{aligned}
 \sum_{r=1}^5 \left\{ \left[ L_{k_1}, L_{r_1} \right] \dot{L}_{r_1} + \left[ L_{k_1}, l_{r_1} \right] \dot{l}_{r_1} \right\} &= - \sum_{p=1}^3 \left( \frac{\partial F_1}{\partial L_{p0}} \frac{\partial L_{p0}}{\partial L_{k_1}} \right. \\
 &\quad \left. + \frac{\partial F_1}{\partial l_{p0}} \frac{\partial l_{p0}}{\partial L_{k_1}} \right) - \sum_{q=4}^5 \left( \frac{\partial F_1}{\partial L_{q0}} \frac{\partial L_{q0}}{\partial L_{k_1}} + \frac{\partial F_1}{\partial l_{q0}} \frac{\partial l_{q0}}{\partial L_{k_1}} \right) \\
 &\quad + \sum_{q=4}^5 \left( \frac{\partial F_2}{\partial L_{q0}} \frac{\partial L_{q0}}{\partial L_{k_1}} + \frac{\partial F_2}{\partial l_{q0}} \frac{\partial l_{q0}}{\partial L_{k_1}} \right)
 \end{aligned} \tag{40}$$

Multiplying each of equations 36a and 36c by  $\partial \ell_{po} / \partial \ell_{k1}$  and  $\partial \ell_{qo} / \partial \ell_{k1}$

respectively and adding gives

$$\begin{aligned} & \sum_{r=1}^5 \left\{ \left[ \sum_{p=1}^3 \frac{\partial L_{po}}{\partial \ell_{r1}} \frac{\partial \ell_{po}}{\partial \ell_{k1}} + \sum_{q=4}^5 \frac{\partial L_{qo}}{\partial \ell_{r1}} \frac{\partial \ell_{qo}}{\partial \ell_{k1}} \right] \dot{L}_{r1} + \left[ \sum_{p=1}^3 \frac{\partial L_{po}}{\partial \ell_{r1}} \frac{\partial \ell_{po}}{\partial \ell_{k1}} \right. \right. \\ & \left. \left. + \sum_{q=4}^5 \frac{\partial L_{qo}}{\partial \ell_{r1}} \frac{\partial \ell_{qo}}{\partial \ell_{k1}} \right] \dot{\ell}_{r1} \right\} = \sum_{p=1}^3 \frac{\partial F_1}{\partial \ell_{po}} \frac{\partial \ell_{po}}{\partial \ell_{k1}} + \sum_{q=4}^5 \frac{\partial F_1}{\partial \ell_{qo}} \frac{\partial \ell_{qo}}{\partial \ell_{k1}} - \sum_{q=4}^5 \frac{\partial F_2}{\partial \ell_{qo}} \frac{\partial \ell_{qo}}{\partial \ell_{k1}} \end{aligned} \quad (41)$$

Multiplying each of equations 36b and 36d by  $\frac{\partial L_{po}}{\partial \ell_{k1}}$  and  $\frac{\partial L_{qo}}{\partial \ell_{k1}}$  respectively and adding gives

$$\begin{aligned} & \sum_{r=1}^5 \left\{ \left[ \sum_{p=1}^3 \frac{\partial \ell_{po}}{\partial \ell_{r1}} \frac{\partial L_{po}}{\partial \ell_{k1}} + \sum_{q=4}^5 \frac{\partial \ell_{qo}}{\partial \ell_{r1}} \frac{\partial L_{qo}}{\partial \ell_{k1}} \right] \dot{L}_{r1} + \left[ \sum_{p=1}^3 \frac{\partial \ell_{po}}{\partial \ell_{r1}} \frac{\partial L_{po}}{\partial \ell_{k1}} \right. \right. \\ & \left. \left. + \sum_{q=4}^5 \frac{\partial \ell_{qo}}{\partial \ell_{r1}} \frac{\partial L_{qo}}{\partial \ell_{k1}} \right] \dot{\ell}_{r1} \right\} = \sum_{p=1}^3 \frac{-\partial F_1}{\partial L_{po}} \frac{\partial L_{po}}{\partial \ell_{k1}} + \sum_{q=4}^5 \frac{-\partial F_1}{\partial L_{qo}} \frac{\partial L_{qo}}{\partial \ell_{k1}} \\ & + \sum_{q=4}^5 \frac{\partial F_2}{\partial L_{qo}} \frac{\partial L_{qo}}{\partial \ell_{k1}} \end{aligned} \quad (42)$$

subtracting equation 41 from equation 42 then gives

$$\begin{aligned} & \sum_{r=1}^5 \left\{ \left[ \ell_{k1}, L_{r1} \right] \dot{L}_{r1} + \left[ \ell_{k1}, \ell_{r1} \right] \dot{\ell}_{r1} \right\} = - \sum_{p=1}^3 \left( \frac{\partial F_1}{\partial L_{po}} \frac{\partial L_{po}}{\partial \ell_{k1}} + \frac{\partial F_1}{\partial \ell_{po}} \frac{\partial \ell_{po}}{\partial \ell_{k1}} \right) \\ & - \sum_{q=4}^5 \left( \frac{\partial F_1}{\partial L_{qo}} \frac{\partial L_{qo}}{\partial \ell_{k1}} + \frac{\partial F_1}{\partial \ell_{qo}} \frac{\partial \ell_{qo}}{\partial \ell_{k1}} \right) + \sum_{q=4}^5 \left( \frac{\partial F_2}{\partial L_{qo}} \frac{\partial L_{qo}}{\partial \ell_{k1}} + \frac{\partial F_2}{\partial \ell_{qo}} \frac{\partial \ell_{qo}}{\partial \ell_{k1}} \right) \end{aligned} \quad (43)$$

To evaluate the brackets it is necessary to transform the determining

function by means of the transform equations (35)



$$S(L_{j1}, l_{j0}) = S'(L_{j1}, l_{j1}) \quad j = 1 - - - 5 \quad (44)$$

The partial derivatives of  $S'$  may then be expressed

$$\begin{aligned} \frac{\partial S'}{\partial L_{k1}} &= l_{k1} + \sum_{j=1}^5 \frac{\partial S}{\partial l_{j0}} \frac{\partial l_{j0}}{\partial L_{k1}} = l_{k1} + \sum_{j=1}^5 L_{j0} \frac{\partial l_{j0}}{\partial L_{k1}} \\ \frac{\partial S'}{\partial l_{k1}} &= \sum_{j=1}^5 \frac{\partial S}{\partial l_{j0}} \frac{\partial l_{j0}}{\partial l_{k1}} = \sum_{j=1}^5 \frac{\partial l_{j0}}{\partial l_{k1}} L_{j0} \end{aligned} \quad (45)$$

rewriting the brackets as

$$\begin{aligned} [L_{k1}, L_{r1}] &= \frac{\partial}{\partial L_{k1}} \left[ \sum_{p=1}^3 L_{p0} \frac{\partial l_{p0}}{\partial L_{r1}} + \sum_{q=4}^5 L_{q0} \frac{\partial l_{q0}}{\partial L_{r1}} \right] \\ &- \frac{\partial}{\partial L_{r1}} \left[ \sum_{p=1}^3 L_{p0} \frac{\partial l_{p0}}{\partial L_{k1}} + \sum_{q=4}^5 L_{q0} \frac{\partial l_{q0}}{\partial L_{k1}} \right] \\ [L_{k1}, l_{r1}] &= \frac{\partial}{\partial L_{k1}} \left[ \sum_{p=1}^3 L_{p0} \frac{\partial l_{p0}}{\partial l_{r1}} + \sum_{q=4}^5 L_{q0} \frac{\partial l_{q0}}{\partial l_{r1}} \right] \\ &- \frac{\partial}{\partial l_{r1}} \left[ \sum_{p=1}^3 L_{p0} \frac{\partial l_{p0}}{\partial L_{k1}} + \sum_{q=4}^5 L_{q0} \frac{\partial l_{q0}}{\partial L_{k1}} \right] \\ [l_{k1}, l_{r1}] &= \frac{\partial}{\partial l_{k1}} \left[ \sum_{p=1}^3 L_{p0} \frac{\partial l_{p0}}{\partial l_{r1}} + \sum_{q=4}^5 L_{q0} \frac{\partial l_{q0}}{\partial l_{r1}} \right] \\ &- \frac{\partial}{\partial l_{r1}} \left[ \sum_{p=1}^3 L_{p0} \frac{\partial l_{p0}}{\partial l_{k1}} + \sum_{q=4}^5 L_{q0} \frac{\partial l_{q0}}{\partial l_{k1}} \right] \end{aligned}$$

and substituting the derivatives of  $S'$

$$[L_{k1}, L_{r1}] = \frac{\partial}{\partial L_{k1}} \left[ \frac{\partial S'}{\partial L_{r1}} - l_{r1} \right] - \frac{\partial}{\partial L_{r1}} \left[ \frac{\partial S'}{\partial L_{k1}} - l_{k1} \right] \quad (46a)$$

$$\left[ \ell_{k_1}, \ell_{r_1} \right] = \frac{\partial}{\partial \ell_{k_1}} \left[ \frac{\partial S'}{\partial \ell_{r_1}} \right] - \frac{\partial}{\partial \ell_{r_1}} \left[ \frac{\partial S'}{\partial \ell_{k_1}} \right] \quad (46b)$$

$$\left[ L_{k_1}, \ell_{r_1} \right] = \frac{\partial}{\partial L_{k_1}} \left[ \frac{\partial S'}{\partial \ell_{r_1}} \right] - \frac{\partial}{\partial \ell_{r_1}} \left[ \frac{\partial S'}{\partial L_{k_1}} - \ell_{k_1} \right] \quad (46c)$$

Then, since all  $L_{j_1}$  and  $\ell_{j_1}$  are independent variables, equation 46a gives

$$\left[ L_{k_1}, L_{r_1} \right] = 0$$

equation 46b gives

$$\left[ \ell_{k_1}, \ell_{r_1} \right] = 0 \quad (47)$$

and equation 46c yields

$$\left[ L_{k_1}, \ell_{r_1} \right] = + \frac{\partial \ell_{k_1}}{\partial \ell_{r_1}} = + \begin{cases} 0 & r \neq k \\ 1 & r = k \end{cases} = \delta_{rk}$$

or

$$\left[ \ell_{k_1}, L_{r_1} \right] = - \delta_{rk}$$

Substitution of equations 47 into the equations 40 and 43 gives.

$$\begin{aligned} \dot{\ell}_{r_1} &= - \sum_{p=1}^3 \left( \frac{\partial F_1}{\partial L_{p_0}} \frac{\partial L_{p_0}}{\partial L_{r_1}} + \frac{\partial F_1}{\partial \ell_{p_0}} \frac{\partial \ell_{p_0}}{\partial L_{r_1}} \right) \\ &\quad - \sum_{q=4}^5 \left( \frac{\partial F_1}{\partial L_{q_0}} \frac{\partial L_{q_0}}{\partial L_{r_1}} + \frac{\partial F_1}{\partial \ell_{q_0}} \frac{\partial \ell_{q_0}}{\partial L_{r_1}} \right) + \sum_{q=4}^5 \left( \frac{\partial F_2}{\partial L_{q_0}} \frac{\partial L_{q_0}}{\partial L_{r_1}} + \frac{\partial F_2}{\partial \ell_{q_0}} \frac{\partial \ell_{q_0}}{\partial L_{r_1}} \right) \\ -\dot{L}_{r_1} &= - \sum_{p=1}^3 \left( \frac{\partial F_1}{\partial L_{p_0}} \frac{\partial L_{p_0}}{\partial \ell_{r_1}} + \frac{\partial F_1}{\partial \ell_{p_0}} \frac{\partial \ell_{p_0}}{\partial \ell_{r_1}} \right) - \sum_{q=4}^5 \left( \frac{\partial F_1}{\partial L_{q_0}} \frac{\partial L_{q_0}}{\partial \ell_{r_1}} + \frac{\partial F_1}{\partial \ell_{q_0}} \frac{\partial \ell_{q_0}}{\partial \ell_{r_1}} \right) \\ &\quad + \sum_{q=4}^5 \left( \frac{\partial F_2}{\partial L_{q_0}} \frac{\partial L_{q_0}}{\partial \ell_{r_1}} + \frac{\partial F_2}{\partial \ell_{q_0}} \frac{\partial \ell_{q_0}}{\partial \ell_{r_1}} \right) \end{aligned} \quad (48)$$

Now, the transformation of the previous section transformed

$$F_1(L_{jo}, \ell_{jo}) = F_1^*(L_{j1}) \quad (49)$$

This of course is a special case of the more general transform

$$F_1(L_{jo}, \ell_{jo}) = F_1^*(L_{j1}, \ell_{j1}) \quad (50)$$

Derivatives of  $F_1^*$  may then be obtained as

$$\frac{\partial F_1^*}{\partial L_{r1}} = \sum_{j=1}^5 \left[ \frac{\partial F_1}{\partial L_{jo}} \frac{\partial L_{jo}}{\partial L_{r1}} + \frac{\partial F_1}{\partial \ell_{jo}} \frac{\partial \ell_{jo}}{\partial L_{r1}} \right] \quad (51)$$

and

$$\frac{\partial F_1^*}{\partial \ell_{r1}} = \sum_{j=1}^5 \left[ \frac{\partial F_1}{\partial L_{jo}} \frac{\partial L_{jo}}{\partial \ell_{r1}} + \frac{\partial F_1}{\partial \ell_{jo}} \frac{\partial \ell_{jo}}{\partial \ell_{r1}} \right]$$

where it is recognized that  $\frac{\partial F_1^*}{\partial \ell_{j1}} = 0$  for the special case of equation 49, but the form of equation 51 is used here to maintain symmetry.

Then combining the summations over  $p = 1$  to 3 and  $q = 4$  to 5 into a single summation over  $j = 1$  to 5 in equation 48 and substitution of equations 51, the following form is obtained.

$$\begin{aligned} \dot{\ell}_{r1} &= - \frac{\partial F_1^*}{\partial L_{r1}} + \sum_{q=4}^5 \left( \frac{\partial F_2}{\partial L_{qo}} \frac{\partial L_{qo}}{\partial L_{r1}} + \frac{\partial F_2}{\partial \ell_{qo}} \frac{\partial \ell_{qo}}{\partial L_{r1}} \right) \\ \dot{L}_{r1} &= + \frac{\partial F_1^*}{\partial \ell_{r1}} - \sum_{q=4}^5 \left( \frac{\partial F_2}{\partial L_{qo}} \frac{\partial L_{qo}}{\partial \ell_{r1}} + \frac{\partial F_2}{\partial \ell_{qo}} \frac{\partial \ell_{qo}}{\partial \ell_{r1}} \right) \end{aligned} \quad (52)$$

The same transformation must now be applied to  $F_2$ . This yields

$$F_2 (L_{j0}, l_{j0}) \Rightarrow F_2' (L_{j1}, l_{j1}) \quad (j=1, \dots, 5) \quad (53)$$

The necessary derivatives from equation 53 are then

$$\begin{aligned} \frac{\partial F_2'}{\partial L_{r1}} &= \sum_{j=1}^5 \left( \frac{\partial F_2}{\partial L_{j0}} \frac{\partial L_{j0}}{\partial L_{r1}} + \frac{\partial F_2}{\partial l_{j0}} \frac{\partial l_{j0}}{\partial L_{r1}} \right) \\ \frac{\partial F_2'}{\partial l_{r1}} &= \sum_{j=1}^5 \left( \frac{\partial F_2}{\partial L_{j0}} \frac{\partial L_{j0}}{\partial l_{r1}} + \frac{\partial F_2}{\partial l_{j0}} \frac{\partial l_{j0}}{\partial l_{r1}} \right) \end{aligned} \quad (54)$$

Now, referring to the transform equations (35) and remembering from the previous section that all  $A_{1i}$  and  $E_{ji}$  are terms of order  $\epsilon_1$  ( $O\epsilon_1$ ), it is seen that the derivatives of the old parameters in terms of the new may be expressed as

$$\frac{\partial L_{j0}}{\partial l_{r1}} = O\epsilon_1 \quad \frac{\partial l_{j0}}{\partial l_{r1}} = \delta_{jr} + O\epsilon_1 \quad (55)$$

$$\frac{\partial L_{j0}}{\partial L_{r1}} = \delta_{jr} + O\epsilon_1 \quad \frac{\partial l_{j0}}{\partial L_{r1}} = O\epsilon_1 \quad \text{where } \delta_{jr} = \begin{cases} 0 & j \neq r \\ 1 & j = r \end{cases}$$

Further, the function  $F_2$  contains a multiplier  $\mu^2 / L_{10}^6$  which is always a small quantity of order less than  $\epsilon_1$  even though it is not a constant. Hence, by neglecting products of these two small quantities equations (54) may be expressed

$$\begin{aligned} \frac{\partial F_2'}{\partial L_{r1}} &= \frac{\partial F_2}{\partial L_{r0}} \\ \frac{\partial F_2'}{\partial l_{r1}} &= \frac{\partial F_2}{\partial l_{r0}} \end{aligned} \quad (r = 1, \dots, 5) \quad (56)$$

Then, substitution of the relations (55) into the equations (52) and again neglecting products of the small quantities, the equations of motion may be expressed

$$\dot{L}_{r1} = \frac{\partial F_1^*}{\partial \dot{l}_{r1}} - \sum_{q=4}^5 \frac{\partial F_2}{\partial \dot{l}_{q0}} \delta_{qr} \quad r = 1, \dots, 5 \quad (57)$$

$$\dot{l}_{r1} = \frac{-\partial F_1^*}{\partial L_{r1}} + \sum_{q=4}^5 \frac{\partial F_2}{\partial L_{q0}} \delta_{qr}$$

Then, substituting equations (56) and taking advantage of the properties of the Kronecker delta, equations (57) may be expressed in expanded form.

$$\begin{aligned} \dot{L}_{11} &= \partial F_1^* / \partial \dot{l}_{11} & \dot{l}_{11} &= - \partial F_1^* / \partial L_{11} \\ \dot{L}_{21} &= \partial F_1^* / \partial \dot{l}_{21} & \dot{l}_{21} &= - \partial F_1^* / \partial L_{21} \\ \dot{L}_{31} &= \partial F_1^* / \partial \dot{l}_{31} & \dot{l}_{31} &= - \partial F_1^* / \partial L_{31} \\ \dot{L}_{41} &= \partial F_1^* / \partial \dot{l}_{41} - \frac{\partial F_2'}{\partial \dot{l}_{41}} & \dot{l}_{41} &= - \partial F_1^* / \partial L_{41} + \frac{\partial F_2'}{\partial L_{41}} \\ \dot{L}_{51} &= \partial F_1^* / \partial \dot{l}_{51} - \frac{\partial F_2'}{\partial \dot{l}_{51}} & \dot{l}_{51} &= - \partial F_1^* / \partial L_{51} + \frac{\partial F_2'}{\partial L_{51}} \end{aligned} \quad (58)$$

Equations (58) are the new equations of motion to be solved. Now note, that  $F_1^*$  of equation 28 contains none of the  $\dot{l}_{j1}$  terms. Hence, the first three equations in the left hand column of equations (58) become:

$$\dot{L}_{11} = 0 \quad ; \quad \dot{L}_{21} = 0 \quad ; \quad \dot{L}_{31} = 0 \quad (59)$$

from whence

$$L_{11} = a_1 = \text{const.}; \quad L_{21} = a_2 = \text{const.}; \quad L_{31} = a_3 = \text{const.} \quad (60)$$

Similarly, the second and third equations in the right hand column of equations (58) become

$$\dot{l}_{21} = 0 ; \quad \dot{l}_{31} = \alpha \quad (61)$$

where  $\alpha$  is a previously defined constant. Hence,

$$l_{21} = c_2 = \text{const.} ; \quad l_{31} = \alpha t + c_3 \quad (62)$$

The first equation in the right hand column of equations (58) becomes

$$\dot{l}_{11} = \frac{\mu^2}{L_{11}^3} - \epsilon_1 \frac{1.7 L_{11}}{\sqrt{1 + (L_{41} / L_{11})^4}} \quad (63)$$

To continue further with this approach, it is necessary that equation 63 take the form

$$\dot{l}_{11} = \beta \quad (64)$$

where  $\beta$  is a constant. The appearance of  $L_{41}$  in the second term of equation 63 thus produces considerable difficulty. Since  $L_{41}$  is related to the Lagrange multipliers, it will in general be unknown. However, it might be noted that if  $L_{41} \gg L_{11}$ , the second term will be much smaller than the first and as such may be neglected. On the other hand, when  $L_{41} \ll L_{11}$  the second term will be of  $O\frac{1}{2}$  as compared with the first term of  $O1$ . Neglection of the second term under these conditions is hardly justified and it will be necessary to assume some constant value for  $L_{41}$  in equation 63.

The procedure from this point on must then be considered iterative in an actual calculation. A good first choice for  $\beta$  will probably be

$$\beta = \mu^2 / a_1^3 \quad (65)$$

where  $a_1$  is the constant of equation (60). The problem must then be solved and the resultant range of  $L_{41}$  and the corresponding range of the neglected term in equation (63) examined. If this neglected term does not remain small compared to  $\mu^2 / a_1^3$  a new  $\beta$  must be chosen

$$\beta = \mu^2 / a_1^3 - \epsilon_1 \cdot 1.7 L_{11} \left[ 1 + (L_{41}' / L_{11})^4 \right]^{-1/2} \quad (66)$$

where  $L_{41}'$  is some averaged constant value from the range of  $L_{41}$  previously calculated. The procedure must then be repeated until the variation in  $L_{41}$  is negligible.

Returning now to the remainder of the problem, with  $l_{11}$  expressed as equation 64,  $l_{11}$  becomes

$$l_{11} = \beta t + c_1 \quad (67)$$

The relations for  $L_{11}, L_{21}, L_{31}, l_{11}, l_{21}, l_{31}$  from equations 60, 62 and 67 may then be substituted into  $F_1^*$  and  $F_2'$  and a new function defined as

$$H' \left[ L_{41}, L_{51}, l_{41}, l_{51}, (a_1, a_2, a_3, \alpha, \beta, c_1, c_2, c_3), t \right]_{\text{after substitution}} = (F_1^* - F_2') \quad (68)$$

and the equations of the problem become

$$\begin{aligned} \dot{L}_{41} &= \partial H' / \partial \ell_{41} & \dot{\ell}_{41} &= - \partial H' / \partial L_{41} \\ \dot{L}_{51} &= \partial H' / \partial \ell_{51} & \dot{\ell}_{51} &= - \partial H' / \partial L_{51} \end{aligned} \quad (69)$$

However, to apply the same procedure as used in the first transform, it is necessary to remove the explicit appearance of time from the Hamiltonian of the problem. To do this, it is necessary to introduce accessory variables and define a new Hamiltonian as

$$H = H' - \beta L_{61} - \alpha L_{71} \quad (70)$$

The equations to be solved are then

$$\begin{aligned} \dot{L}_{41} &= \partial H / \partial \ell_4 & \dot{\ell}_{41} &= - \partial H / \partial L_{41} \\ \dot{L}_{51} &= \partial H / \partial \ell_5 & \dot{\ell}_{51} &= - \partial H / \partial L_{51} \\ \dot{L}_{61} &= \partial H / \partial \ell_6 & \dot{\ell}_{61} &= - \partial H / \partial L_{61} \\ \dot{L}_{71} &= \partial H / \partial \ell_7 & \dot{\ell}_{71} &= - \partial H / \partial L_{71} \end{aligned} \quad (71)$$



## THE SECOND TRANSFORM

The part of the forcing function that was neglected in the first transform may be expressed in the form

$$F_2 = -\frac{\mu^2}{L_{10}^6} \left\{ T_o(L_{jo}) + \sum_{r=1}^n R_{or}(L_{jo}) \cos \psi_{or} \right\} \quad (72)$$

(j = 1, - - - 5)

where

$$\psi_{or} = \sum_{j=1}^5 w_{jr} \ell_{jo}$$

Substitution of the first transform relations of equations (27) yields the terms in  $F_2$  in the following form.

$$T_o(L_{jo}) = T_1(L_{j1}) + \sum_u U_{1u}(L_{j1}) \cos \theta_{1u}$$

$$R_{or}(L_{jo}) = R_{1r}(L_{j1}) + \sum_v V_{1v}(L_{j1}) \cos \theta_{1v}$$

$$L_{10}^{-6} = L_{11}^{-6} (1 - 6 \sum_i p_{1i} \frac{A_{1i}}{L_{11}} \cos \theta_{1i})$$

$$\begin{aligned} \cos \psi_{or} = & \cos \psi_{1r} + \frac{1}{2} \sum_j \sum_i w_{jr} E_{ji} [\cos(\theta_{1i} - \psi_{1r}) \\ & - \cos(\theta_{1i} + \psi_{1r})] \end{aligned}$$

Where  $T_1$  and  $R_{1r}$  are identical expressions to  $T_o$  and  $R_{or}$  with the  $L_{jo}$  replaced by  $L_{j1}$ , and  $\psi_{1r}$  is identical to  $\psi_{or}$  with the  $\ell_{jo}$  replaced by  $\ell_{j1}$ .  $U_{1u}$  and  $V_{1v}$  like  $A_{1i}$  are terms of first order in  $\epsilon_1$ .

Rearranging and reordering the cosine terms,  $F_2$  may be expressed

as

$$F_2' = - \frac{\mu^2}{L_{11}^6} \left[ T_1 + \sum_{h=1}^9 B_{1h} (L_{j_1}) \cos \phi_{1h} \right] \quad (73)$$

where

$$\phi_{1h} = \sum_{j=1}^5 q_{jh} \ell_{j_1}$$

Substituting the solutions of equations 60, 62, and 67, incorporating the auxiliary variables and denoting  $F_2'$ ,  $T_1$  and  $B_1$  after the substitution as  $F_2^*$ ,  $T_1^*$ ,  $B_1^*$ , and  $\mu^2/a_1^6$  as  $\epsilon_2$ , the expression becomes

$$F_2^* = - \epsilon_2 \left[ T_1^* (L_{41}, L_{51}, a_1, a_2) + \sum_{k=1}^m B_{1k}^* (L_{41}, L_{51}, a_1, a_2) \cos \phi_{1k} \right] \quad (74)$$

where

$$\phi_{1k} = \sum_{i=4}^7 q_{ik} \ell_{i_1} + q_{2k} c_2$$

Performing the substitutions in  $F_1^*$  and adding the auxiliary terms, the Hamiltonian for the problem becomes

$$H = \frac{\mu^2}{2a_1^2} - \frac{\mu^2}{2L_{41}^2} - \alpha a_3 - .85\epsilon_1 (a_1^4 + L_{41}^4)^{1/2} - \beta L_{61} - \alpha L_{71} + \epsilon_2 \left[ T_1^* + \sum_{k=1}^m B_{1k}^* \cos \phi_{1k} \right] \quad (75)$$

The equations are given as equations (71) which may be expressed

$$\dot{L}_{j_1} = \partial H / \partial \ell_{j_1} \quad \dot{\ell}_{j_1} = - \partial H / \partial L_{j_1} \quad (j = 4, \dots, 7) \quad (76)$$

The second transform now follows the same procedure as the first transform.  $\epsilon_2$  is the small parameter and  $H$  may be split into two parts,  $H_0$  void of  $\epsilon_2$ , and  $H_1$  all terms of which contain  $\epsilon_2$ .

$$H_0 = -\frac{\mu^2}{2L_{41}^2} - .85 \epsilon_1 (a_1^4 + L_{41}^4)^{1/2} - \beta L_{61} - \alpha L_{71} \quad (77a)$$

where the terms  $\frac{\mu^2}{2a_1^2} - \alpha a_3$  have been neglected since they don't affect the solution.

$$H_1 = X_1 + \sum_{k=1}^m Y_{1k} \cos \phi_{1k} \quad (77b)$$

where

$$X_1 = \epsilon_2 T_1^* \text{ and } Y_{1k} = \epsilon_2 B_{1k}^*$$

The object now is to transform the function  $H(L_{j1}, \ell_{j1})$  into a function of the  $L_{j2}$  only

$$H(L_{j1}, \ell_{j1}) = H^*(L_{j2}) \quad (j = 4, \dots, 7) \quad (78)$$

As in the first transform, a determining function will again be used in the form

$$S = S(L_{j2}, \ell_{j1})$$

which gives the transform relations

$$L_{j1} = \partial S / \partial \ell_{j1} \quad \ell_{j2} = \partial S / \partial L_{j2} \quad (j = 4, \dots, 7) \quad (79)$$

The parts of the Hamiltonian in equations (77) then appear in functional form upon substitution of equations (79) as

$$H_0 = H_0(\partial S / \partial \ell_{j1}) \quad H_1 = H_1\left(\frac{\partial S}{\partial \ell_{j1}}, \ell_{j1}\right) \quad (j = 4, \dots, 7) \quad (80)$$

The determining function may then be expanded in powers of the small parameter  $\epsilon_2$ .

$$S = S_0 + S_1 + S_2 + \dots \quad (81)$$

Where again, to insure the identity transformation when  $\epsilon_2$  is zero,

$S_0$  is taken in the form

$$S_0 = \sum_{j=4}^7 L_{j2} \ell_{j1} \quad (82)$$

Expanding the new Hamiltonian  $H^*(L_{j2})$  in powers of the small parameter  $\epsilon_2$  and substituting equations (82) and (80) into (78), the Hamilton-Jacobi equation becomes

$$\begin{aligned} H_0 \left[ L_{j2} + \left( \frac{\partial S_1}{\partial \ell_{j1}} + \frac{\partial S_2}{\partial \ell_{j1}} + \dots \right) \right] + H_1 \left[ L_{j2} + \left( \frac{\partial S_1}{\partial \ell_{j1}} + \frac{\partial S_2}{\partial \ell_{j1}} + \dots \right), \right. \\ \left. \phi_{1k} \right] = H_0^* + H_1^* + H_2^* + \dots \quad (83) \end{aligned}$$

$H_0$  and  $H_1$  may then be expanded Taylor's series which to the first order in  $t_2$  become

$$\begin{aligned} H_0 = H_0(L_{j2}) + \sum_{j=4}^7 \left. \frac{\partial H_0}{\partial L_{j1}} \right|_{L_{j2}} \left( \frac{\partial S_1}{\partial \ell_{j1}} \right) \\ H_1 = H_1(L_{j2}, \phi_{1k}) \quad (84) \end{aligned}$$

Substituting equations (84) into (83) and equating terms of like order in  $\epsilon_2$  yields

$$H_0(L_{j2}) = H_0^* \quad (85a)$$

$$\begin{aligned} \sum_{j=4}^7 \left. \frac{\partial H_0}{\partial L_{j1}} \right|_{L_{j2}} \frac{\partial S_1}{\partial \ell_{j1}} + X_2(L_{j2}) + \sum_k Y_{2k}(L_{j2}) \cos \phi_{1k} \\ = H_1^*(L_{j2}) \quad (85b) \end{aligned}$$

where  $X_2$  and  $Y_{2k}$  are identical expressions to  $X_1$  and  $Y_{1k}$  with  $L_{j1}$  replaced by  $L_{j2}$ .

The necessary derivatives are evaluated from equations (77) as

$$\begin{aligned}
 \left. \frac{\partial H_o}{\partial L_{41}} \right|_{L_{j2}} &= + \frac{\mu^2}{L_{42}^3} - 1.7\epsilon_1 (a_1^4 + L_{42}^4)^{-1/2} L_{42}^3 = -n_{42} \\
 \left. \frac{\partial H_o}{\partial L_{51}} \right|_{L_{j2}} &= 0 = -n_{52} \\
 \left. \frac{\partial H_o}{\partial L_{61}} \right|_{L_{j2}} &= -\beta = -n_{62} \\
 \left. \frac{\partial H_o}{\partial L_{71}} \right|_{L_{j2}} &= -\alpha = -n_{72}
 \end{aligned} \tag{86}$$

Using the notation of  $n_{j2}$  as given in equation (86), the derivatives in summation form may be expressed

$$\sum_{j=4}^7 \left. \frac{\partial H_o}{\partial L_{j1}} \right|_{L_{j2}} \frac{\partial S_1}{\partial l_{j1}} = \sum_{j=4}^7 -n_{j2} \frac{\partial S_1}{\partial l_{j1}} \tag{87}$$

Referring now to equation (85b), it is seen that the second term  $X_2$  and the right hand side,  $H_1^*$ , are both functions of the  $L_{j2}$  only while the other two terms are functions of both the  $L_{j2}$  and the  $l_{j1}$ . Therefore, the  $H_1^*$  is only related to  $X_2$ .

$$H_1^* (L_{j2}) = X_2 (L_{j2}) \tag{88}$$

Substitution of equations (87) and (88) in (85b) then yields

$$\sum_{j=4}^7 n_{j2} \frac{\partial S_1}{\partial \ell_{j1}} = - \sum_k Y_{2k} (L_{j2}) \cos \phi_{1k} \quad (89)$$

A solution of equation (89) may be taken as

$$S_1 = \sum_k C_{2k} (L_{j2}) \sin \phi_{1k} \quad (90)$$

Substitution of equation (90) into (89) then gives

$$\sum_k \sum_{j=4}^7 C_{2k} q_{jk} n_{j2} \cos \phi_{1k} = \sum_k Y_{2k} \cos \phi_{1k} \quad (91)$$

from whence

$$C_{2k} = \frac{Y_{2k}}{\sum_{j=4}^7 q_{jk} n_{j2}} \quad (92)$$

Equations (81), (82), and (90) then give the first order determining function

$$S = \sum_{j=4}^7 L_{j2} \ell_{j1} + \sum_k C_{2k} \sin \phi_{1k} \quad (93)$$

Substitution of equation (93) into the transform relations (79) then gives the transform equations

$$\begin{aligned} L_{j1} &= L_{j2} + \sum_k C_{2k} q_{jk} \cos \phi_{1k} \\ \ell_{j2} &= \ell_{j1} + \sum_k D_{jk} \sin \phi_{1k} \end{aligned} \quad (j = 4, \dots, 7) \quad (94)$$

where

$$D_{jk} = \frac{\partial C_{2k}}{\partial L_{j2}}$$

Substitution into the expression for  $\phi_{1k}$  and performing Lagrange expansions to the first order in  $\epsilon_2$  (as was done in the first transform), the transform equations become

$$\begin{aligned} L_{j1} &= L_{j2} + \sum_k C_{2k} q_{jk} \cos \phi_{2k} \\ \ell_{j1} &= \ell_{j2} - \sum_k D_{jk} \sin \phi_{2k} \end{aligned} \quad (j = 4, \dots, 7) \quad (95)$$

where

$$\phi_{2k} = \sum_{j=4}^7 q_{jk} \ell_{j2} + q_{2k} c_2 \quad (96)$$

The complete first order Hamiltonian in the new variables may then be obtained from equations (77a), (85a) and (88) as

$$H^*(L_{j2}) = -\frac{\mu^2}{2L_{42}^2} - .85\epsilon_1 (a_1^4 + L_{42}^4)^{1/2} - \beta L_{62} - \alpha L_{72} + \epsilon_2 T_2(L_{j2}) \quad (97)$$

where the expression  $X_2 = \epsilon_2 T_2$  has been incorporated.

The new equations of motion to be solved are

$$L_{j2} = \frac{\partial H^*}{\partial \ell_{j2}} \quad \ell_{j2} = -\partial H^* / \partial L_{j2} \quad (j = 4, \dots, 7) \quad (98)$$

These equations have solutions of the form

$$\begin{aligned} L_{j2} &= a_j & (j = 4, \dots, 7) \\ \ell_{j2} &= b_j t + c_j & (j = 4, \dots, 7) \end{aligned} \quad (99)$$

Where  $a_j$  and  $c_j$  are constants as well as the  $b_j$  which are functions of the  $a_j$  and previously defined constants.  $T_2$  is obtained from the expansion of  $F_2$  in a previous section by incorporating the results of the first transform and then replacing  $L_{41}$  and  $L_{51}$  by  $L_{42}$  and  $L_{52}$ .

$$T_2 = \frac{1}{2} L_{42}^4 + a_1^6 L_{42}^{-2} - \frac{3}{64} (a_1 + a_2)^2 a_1^{-2} L_{42}^2 (L_{42} + L_{52})^2 \quad (100)$$

The expressions for  $b_j$  may then be obtained from equations (97), (98) and (100) as:

$$\begin{aligned} b_4 &= -\frac{\mu^2}{a_4^3} + \frac{1.7\epsilon_1 a_4^3}{(a_1^4 + a_4^4)^{1/2}} - \epsilon_2 a_4 \left\{ 2a_4^2 (1 - a_1^6 a_4^{-6}) \right. \\ &\quad \left. - \frac{3}{32} \left(1 + \frac{a_2}{a_1}\right)^2 (a_4 + a_5) (2a_4 + a_5) \right\} \\ b_5 &= + \frac{3}{32} \epsilon_2 a_4^2 \left(1 + \frac{a_2}{a_1}\right)^2 (a_4 + a_5) \\ b_6 &= \beta \\ b_7 &= \alpha \end{aligned} \quad (101)$$



## THE SOLUTION FORM

The preceeding sections have explained the transformations necessary to integrate the differential equations (5) or (5a). To be useful, it is necessary to express the original variables ( $L_{j0}, \ell_{j0}$ ) in terms of the constants of integration.

The two transform relations were obtained as equations (27) and (95) and are repeated here for compactness of the following discussion.

$$\begin{aligned} L_{j0} &= L_{j1} + \sum_i A_{1i} (L_{j1}) p_{ji} \cos \phi_{1i} \\ \ell_{j0} &= \ell_{j1} - \sum_i E_{ji} (L_{j1}) \sin \phi_{1i} \end{aligned} \quad (j = 1, \dots, 5) \quad (27)$$

and.

$$\begin{aligned} L_{j1} &= L_{j2} + \sum_k C_{2k} (L_{j2}) q_{jk} \cos \phi_{2k} \\ \ell_{j1} &= \ell_{j2} - \sum_k D_{jk} (L_{j2}) \sin \phi_{2k} \end{aligned} \quad (j = 4, \dots, 7) \quad (95)$$

Also, in the process of determining a canonical form for the second transform, the following relations were generated

$$L_{j1} = a_j \quad (j = 1, \dots, 3) \quad (102)$$

$$\ell_{11} = \beta t + c_1; \quad \ell_{21} = c_2; \quad \ell_{31} = \alpha t + c_3 \quad (103)$$

Now, denoting  $a_j = L_{j_2}$  ( $j = 1, \dots, 3$ ) to aid in notation, the equations

$$L_{j_1} = L_{j_2} \quad (j = 1, \dots, 3) \quad (104)$$

may be considered as part of the second transform equations.

Next, each  $A_{1i}(L_{j_1})$  may be expanded in a Taylor's series about the point  $L_{j_1} = L_{j_2}$  to the first order

$$A_{1i}(L_{j_1}) = A_{2i}(L_{j_2}) + \sum_{j=1}^5 \frac{\partial A_{1i}}{\partial L_{j_1}} \bigg|_{L_{j_2}} (L_{j_1} - L_{j_2}) + \dots \quad (105)$$

The terms  $(L_{j_1} - L_{j_2})$  are determined from equations (104) and (95) as

$$\begin{aligned} (L_{j_1} - L_{j_2}) &= 0 \quad (j = 1, \dots, 3) \\ (L_{j_1} - L_{j_2}) &= \sum_k C_{2k} q_{jk} \cos \phi_{2k} \quad (j = 4, 5) \end{aligned} \quad (106)$$

Hence

$$A_{1i} = A_{2i}(L_{j_2}) + \sum_{j=4}^5 \sum_k \frac{\partial A_{2i}}{\partial L_{j_2}} C_{2k} q_{jk} \cos \phi_{2k} \quad (107)$$

where  $A_{2i}$  is an identical expression to  $A_{1i}$  with the  $L_{j_1}$  replaced by  $L_{j_2}$

Equations (107) and the first of (95) may then be substituted into the first of equations (27) to yield

$$\begin{aligned} L_{j_0} &= L_{j_2} + \sum_k C_{2k} q_{jk} \delta_j \cos \phi_{2k} + \sum_i A_{2i} p_{ji} \cos \theta_{1i} \\ &+ \sum_i \left\{ \sum_{h=4}^5 \sum_k \frac{\partial A_{2i}}{\partial L_{h_2}} C_{2k} q_{hk} \cos \phi_{2k} p_{ji} \cos \theta_{1i} \right\} \end{aligned}$$

Now, noting that  $A_{2i}$  is of order  $\epsilon_1$  and  $C_{2k}$  is of order  $\epsilon_2$  and neglecting terms of order  $\epsilon_1\epsilon_2$ , this becomes

$$L_{j0} = L_{j2} + \sum_k C_{2k} q_{jk} \delta_j \cos \phi_{2k} + \sum_i A_{2i} p_{ji} \cos \theta_{1i} \quad (108)$$

where

$$\delta_j = \begin{cases} 0 & j = 1, 2, 3 \\ 1 & j = 4, 5 \end{cases}$$

Next, from equations (99) and (101) it is seen that  $\ell_{62} = \ell_{61}$  and  $\ell_{72} = \ell_{71}$ . Also, in the formulation of the canonical form it was specified that

$$\ell_{61} = \beta t + c_1 = \ell_1$$

and

$$\ell_{71} = \alpha t + c_3 = \ell_{31}$$

and in addition,  $q_{6k} = q_{1k}$  and  $q_{7k} = q_{3k}$ . Consequently, it may be specified that

$$\begin{aligned} \ell_{11} &= \ell_{12} \\ \ell_{21} &= \ell_{22} \\ \ell_{31} &= \ell_{32} \end{aligned} \quad (108)$$

are transform relations replacing the  $\ell_{61}$  and  $\ell_{71}$  relations of equations (95). With this notation,  $\phi_{2k}$  becomes

$$\phi_{2k} = \sum_{i=1}^5 q_{ik} \ell_{i2} \quad (109)$$

Now, returning to equation (108),  $\theta_{1i}$  may be expressed

$$\theta_{1i} = \sum_{j=1}^5 p_{ji} \ell_{j1}$$

substituting for the  $\ell_{j1}$  from equations (109) and (95) gives

$$\theta_{1i} = \sum_{j=1}^5 p_{ji} \ell_{j2} - \sum_{j=4}^5 p_{ji} \sum_k D_{jk} \sin \phi_{2k}$$

Defining

$$\theta_{2i} = \sum_{j=1}^5 p_{ji} \ell_{j2}$$

$$\theta_{1i} = \theta_{2i} - \sum_{j=4}^5 \sum_k D_{jk} p_{ji} \sin \phi_{2k} \quad (110)$$

and

$$\begin{aligned} \cos \theta_{1i} = \cos \theta_{2i} + \frac{1}{2} \sum_{j=4}^5 \sum_k D_{jk} p_{ji} \left[ \cos (\phi_{2k} - \theta_{2i}) - \cos (\phi_{2k} + \theta_{2i}) \right] \\ + (\text{terms of higher order in } \epsilon_2) \end{aligned} \quad (111)$$

Substitution of equation (111) in equation (108) then gives

$$\begin{aligned} L_{jo} = L_{j2} + \sum_i A_{zi} p_{ji} \cos \theta_{2i} + \sum_k C_{2k} q_{jk} \delta_j \cos \phi_{2k} \\ + \frac{1}{2} \sum_i \sum_{h=4}^5 \sum_k \sum_l A_{zi} p_{ji} D_{hk} p_{hi} \left[ \cos (\phi_{2k} - \theta_{2i}) - \cos (\phi_{2k} + \theta_{2i}) \right] \end{aligned}$$

or again neglecting terms of order  $\epsilon_1 \epsilon_2$ ,

$$L_{jo} = L_{j2} + \sum_i A_{zi} p_{ji} \cos \theta_{2i} + \sum_k C_{2k} q_{jk} \delta_j \cos \phi_{2k} \quad (112)$$

Now returning to equation (27), the  $E_{ji}(L_{j1})$  may be expanded in Taylor's series in the same manner as the  $A_{1i}$ .

$$E_{ji}(L_{j1}) = E_{ji}(L_{j2}) + \sum_{h=4}^5 \sum_k \frac{\partial E_{ji}}{\partial L_{j2}} C_{2k} q_{hk} \cos \phi_{2k} \quad (113)$$

Substituting equation (113) along with equations (108) and the second of equations (95) into the second of equations (27) then yields

$$\begin{aligned} \ell_{jo} = \ell_{j2} - \sum_k D_{jk} \delta_j \sin \phi_{2k} - \sum_i E_{ji}(L_{j2}) \sin \theta_{1i} \\ - \sum_i \sum_{h=4}^5 \sum_k \frac{\partial E_{ji}(L_{j2})}{\partial L_{j2}} C_{2k} q_{hk} \cos \phi_{2k} \sin \theta_{1i} \end{aligned} \quad (114)$$

From equation (110) neglecting terms of order  $\epsilon_2^2$  and higher,

$$\begin{aligned} \sin \theta_{1i} = \sin \theta_{2i} - \frac{1}{2} \sum_{j=4}^5 \sum_k D_{jk} p_{ji} \left[ \sin(\phi_{2k} + \theta_{2i}) \right. \\ \left. + \sin(\phi_{2k} - \theta_{2i}) \right] \end{aligned} \quad (115)$$

Substitution of equation (115) into (114) and neglecting terms of order  $\epsilon_1 \epsilon_2$  and higher gives

$$\ell_{jo} = \ell_{j2} - \sum_k D_{jk} \delta_j \sin \phi_{2k} - \sum_i E_{ji}(L_{j2}) \sin \theta_{2i} \quad (116)$$

Replacing the  $L_{j2}$  and  $\ell_{j2}$  terms in equations (112) and (116) by their constant values then gives the relations between the original variables, the constants of integration and time.

$$\begin{aligned} L_{jo} = a_j + \sum_i A_{2i}(a_h) p_{ji} \cos \theta_{2i}(b_h t + c_h) + \sum_k C_{2k}(a_h) q_{jk} \delta_j \cos \phi_{2k}(b_h t + c_h) \\ \ell_{jo} = (b_j t + c_j) - \sum_i E_{ji}(a_h) \sin \theta_{2i}(b_h t + c_h) - \sum_k D_{jk}(a_h) \delta_j \sin \phi_{2k}(b_h t + c_h) \end{aligned} \quad (117)$$

(j = 1, ---, 5) (h = 1, ---, 5)

where

$$\delta_j = \begin{cases} 0 & j = 1, 2, 3 \\ 1 & j = 4, 5 \end{cases}$$

and

$$b_1 = \beta; \quad b_2 = 0; \quad b_3 = \alpha$$

$b_4$  and  $b_5$  are given in equation (101) and the constants  $a_1, \dots, a_5$ ,

$c_1, \dots, c_5$  are the constants of integration obtained previously.

## CONCLUSIONS

The analytical procedure for obtaining a first order approximate solution to equations evolving from equations representing a general minimum fuel low thrust problem has been presented. The actual evaluation of the constants of integration depends of course on the nature of each particular problem.

It may be expected that this procedure, especially in the first order format presented here, will be more applicable to situations in which the vehicle makes many orbits around a central body to attain orbital transfer or to rendezvous with or intercept another orbital vehicle. Calculations of interplanetary transfer with this procedure will probably require higher order approximations.

The determination of higher order approximations with this procedure is straight forward through the first transform. This requires the simple (though tedious) expansion of the  $F_1$  disturbing function to higher powers of the eccentricity; the inclusion of higher order terms in the expansions of the determining function,  $F_1^*$  and all Taylor's series; and the performance of the extra steps required to determine the higher order terms of the determining function. The extensions required in obtaining a canonical form for the second transform are not obvious. However, it might be noted that the need for the higher

order solution will most likely result when the conical eccentricities are expected to be large. A look at the term  $\epsilon_2$  which multiplies the entire  $F_2$  term indicates that it is a function of the inverse cube of the semi-major axis. Hence, as the eccentricity increases  $\epsilon_2$  decreases at a much faster rate and the neglect of the entire  $F_2$  term may be feasible.

The next step in the development of this procedure will be the attainment of numerical results for a physical problem. The weakest point of the method so far (other than the order limitation) appears to be the iteration procedure that is required to obtain a good value for the  $\beta$  term in obtaining the canonical formulation for the second transform.



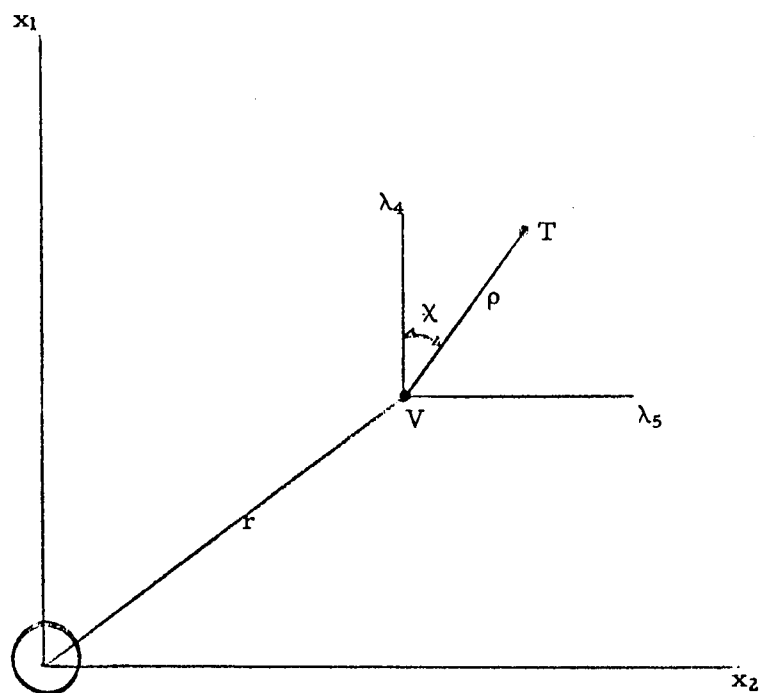


FIGURE 1

## REFERENCES

1. Cox, J. G. and Shaw, W. A., "Preliminary Investigations on Three Dimensional Optimum Trajectories", Progress Report No. 1 on Studies in the Fields of Space Flight and Guidance Theory. NASA Report MTP-AERO-61-91, dated 18 December 1961.
2. Brouwer, Dick and Clemence, G. M., "Methods of Celestial Mechanics". Academic Press, New York, 1961.
3. Smart, W. M., "Celestial Mechanics". John Wiley and Sons, Inc., New York, 1953.

RESEARCH DEPARTMENT  
GRUMMAN AIRCRAFT ENGINEERING CORPORATION  
BETHPAGE, NEW YORK

---

SOME NUMERICAL RESULTS OF GEOCENTRIC  
LOW THRUST TRAJECTORY OPTIMIZATION

by

Hans K. Hinz  
Robert McGill  
Gerald Taylor

SUMMARY

The generalized Newton-Raphson method is used to determine optimum, coplanar, circle-to-circle, transfer trajectories for low thrust space vehicles operating in a strong central force field, such as a near earth orbit. Optimum thrust steering programs are computed for progressively increasing values of final time up to durations involving 26 revolutions about the earth. A description of the numerical results and a comparison of these with the results of a previous linear analysis are given.

*Author*

33061  
M65 33061

## INTRODUCTION

This paper is concerned with the computation of optimum, orbital transfer trajectories for space vehicles with low thrust, electrical propulsion systems operating in a strong central force field, such as near-earth orbits. Although the magnitude of thrust acceleration for interplanetary and geocentric low thrust missions can be similar, the optimum trajectories for the two missions are quite different. This is due to the predominant gravitational attraction of the earth, which at an altitude of 200 miles is more than 1500 times greater than the gravitational attraction of the sun at a distance of one astronomical unit. Thus, many orbital circuits are required for a low thrust vehicle to complete various geocentric missions.

Many of the problems associated with optimization of geocentric low thrust trajectories stem from the large number of revolutions about the earth required of the vehicle. One of these problems is the sizable accumulation of round-off and truncation error resulting from the many integration intervals. A second difficulty, associated with some of the successive approximation techniques, is the need to store the control variables as functions of time. If the functions are rapidly changing ones, the amount of computer storage required may become prohibitive. A third difficulty, usually associated with the classical indirect methods for solving the boundary value problem, is the extreme sensitivity of terminal conditions to initial conditions of the multipliers. As the number of revolutions for an optimum trajectory increases, the sensitivity may be intensified to a point where systematic computer procedures will not converge to the desired solution.

Several successive approximation techniques, each employing a variation-of-parameters integration procedure, have been developed (Refs. 1 and 2) and programmed for IBM 7090 computation. Although these methods have proven partially successful, satisfactory convergence to solutions of the multiple pass problem have not been achieved. As an alternate approach to this problem, the generalized Newton-Raphson method (Refs. 3 and 4) has been used with consider-

able success. The algorithm for this method solves a sequence of linear boundary value problems such that the sequence of solutions converges to the solution of the non-linear problem. Because the linear boundary value problem is easily handled numerically, the algorithm is readily adaptable to high speed, digital computation. Another advantage is that the initial approximations do not have to satisfy the differential equations or the boundary conditions. Thus, simple starting functions, such as straight lines or unperturbed two-body orbits, are usually adequate for convergence to the desired solution.

The specific problem treated in this paper is that of determining the optimum thrust steering program that will minimize the time to transfer between coplanar, circular orbits. Since the thrust magnitude is fixed, minimum time is equivalent to minimum fuel.

### SYSTEM MODEL

For the system model, only coplanar motion in a geocentric inverse-square gravity field is considered. The vehicle is taken as a mass particle with a thrust vector constant in magnitude and variable in direction. The problem is to determine the optimum thrust steering program for minimum time transfer from a circular orbit at an altitude of 200 statute miles to a higher energy circular orbit. Because the vehicle's mass decreases linearly with time, minimizing time is equivalent to minimizing fuel.

The equations of motion in polar coordinates are

$$\dot{u} = \frac{v^2}{r} - \frac{k}{r^2} + \frac{T \sin \theta}{m_0 + \dot{m}t},$$

$$\dot{v} = -\frac{uv}{r} + \frac{T \cos \theta}{m_0 + \dot{m}t},$$

$$\dot{r} = u,$$

$$\dot{\phi} = \frac{v}{r}.$$

Here,  $k$  is the gravitational parameter of the earth;  $T$  is the thrust;  $\theta$  is the thrust steering angle;  $m_0$  is the initial mass; and  $\dot{m}$  is the time rate of change of the vehicle's mass. The state variables are defined in Fig. 1.

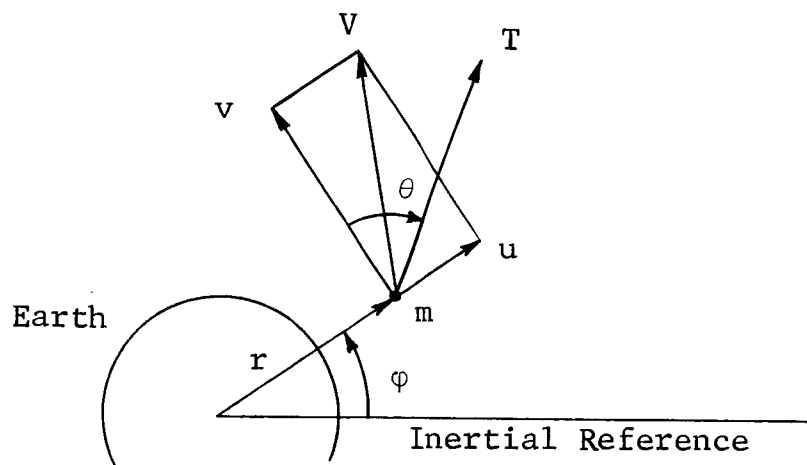


Fig. 1 Coordinate System

The basic numerical data used for most of the computations are

$$k = 1.408 \times 10^{16} \text{ ft}^3/\text{sec}^2 ,$$

$$r_o = 2.19825 \times 10^7 \text{ ft} ,$$

$$g = 32.2 \text{ ft/sec}^2 ,$$

$$T = 10 \text{ lb} ,$$

$$W_o = 10,000 \text{ lb} ,$$

$$I_s = 5000 \text{ sec} ,$$

$$m_o = W_o/g = 310.559 \text{ slug} ,$$

$$\dot{m} = -T/I_s g = -6.21118 \times 10^{-5} \text{ slug/sec} ,$$

$$u_o = 0$$

$$v_o = (k/r_o)^{\frac{1}{2}}$$

where  $g$  is the gravitational acceleration at the surface of the earth,  $W_o$  is the vehicle's initial weight, and  $I_s$  is the specific impulse

## VARIATIONAL TREATMENT

The results of this paper were obtained by the indirect method of the calculus of variations in conjunction with the generalized Newton-Raphson algorithm (Ref. 4). The existence of a solution to the nonlinear optimal control problem is assumed and the necessary conditions are obtained by the application of the Pontryagin maximum principle (Ref. 5). For the problem treated herein, the necessary conditions may also be obtained by classical procedures (Refs. 6 and 7). These necessary conditions form a nonlinear, two-point, boundary value problem. For the problem of this paper, the relevant boundary value problem is given by the following sixth order system:

$$\dot{r} = u = f^{(1)},$$

$$\dot{u} = \frac{v^2}{r} - \frac{k}{r^2} + \frac{a(t)\lambda_u}{(\lambda_u^2 + \lambda_v^2)^{\frac{1}{2}}} = f^{(2)},$$

$$\dot{v} = -\frac{uv}{r} + \frac{a(t)\lambda_v}{(\lambda_u^2 + \lambda_v^2)^{\frac{1}{2}}} = f^{(3)},$$

$$\dot{\lambda}_r = \left[ \frac{v^2}{r^2} - \frac{2k}{r^3} \right] \lambda_u - \frac{uv}{r^2} \lambda_v = f^{(4)},$$

$$\dot{\lambda}_u = -\lambda_r + \frac{v}{r} \lambda_v = f^{(5)},$$

$$\dot{\lambda}_v = -\frac{2v}{r} \lambda_u + \frac{u}{r} \lambda_v = f^{(6)},$$



where

$$a(t) = \frac{T}{m_0 + \dot{m}t} ,$$

and the boundary conditions are

$$\begin{array}{ll} \text{at } t_0 = 0 & \text{at } t = t_f \text{ (} t_f \text{ unspecified)} \\ r(0) = r_0 , & r(t_f) = r_f , \\ u(0) = u_0 , & u(t_f) = u_f , \\ v(0) = v_0 , & v(t_f) = v_f . \end{array}$$

This may be written as

$$\dot{X} = F(X, t) ,$$

where

$$\begin{aligned} X &= (x^{(1)}, \dots, x^{(6)}) , \\ F &= (f^{(1)}, \dots, f^{(6)}) , \end{aligned}$$

and

$$\begin{aligned} x^{(1)}(t) &= r(t) , & x^{(2)}(t) &= u(t) , & x^{(3)}(t) &= v(t) , \\ x^{(4)}(t) &= \lambda_r(t) , & x^{(5)}(t) &= \lambda_u(t) , & x^{(6)}(t) &= \lambda_v(t) . \end{aligned}$$

The generalized Newton-Raphson algorithm proceeds by solving the following sequence of linear, two-point, boundary value problems:

$$\begin{aligned} X_{n+1} &= J(X_n, t) [X_{n+1}(t) - X_n(t)] + F(X_n, t) , \\ n &= 0, 1, 2, \dots, \end{aligned}$$

where  $J(X, t)$  is the Jacobian matrix of partial derivatives of the  $f^{(i)}$  with respect to the  $x^{(i)}$ ,  $i, j=1, \dots, 6$ . The boundary conditions for every  $n$  are those given above. A starting vector,  $X_0(t)$ , and an estimated final time,  $t_{f_0}$ , are assumed and the sequence of linear boundary value problems is solved numerically by the method described in detail in Ref. 4.

The basic starting vector  $X_0(t)$  is of the following simple form:

$$x_0^{(1)}(t) = r_0(t) = r_0 + \frac{r_f - r_0}{t_{f_0}} t ,$$

$$x_0^{(2)}(t) = u_0(t) \equiv 0 ,$$

$$x_0^{(3)}(t) = v_0(t) = \left[ \frac{k}{r_0(t)} \right]^{\frac{1}{2}} ,$$

$$x_0^{(4)}(t) = \lambda_{r_0}(t) \equiv 1 ,$$

$$x_0^{(5)}(t) = \lambda_{u_0}(t) \equiv \begin{cases} c_1 & \text{for } t \in (0, \frac{1}{2} t_{f_0}) \\ c_2 & \text{for } t \in (\frac{1}{2} t_{f_0}, t_{f_0}) \end{cases} ,$$

$$x_0^{(6)}(t) = \lambda_{v_0}(t) \equiv \begin{cases} c_3 & \text{for } t \in (0, \frac{1}{2} t_{f_0}) \\ c_4 & \text{for } t \in (\frac{1}{2} t_{f_0}, t_{f_0}) \end{cases} .$$

Most of the results described in this report were obtained by first producing a solution using the above simple starting vector,  $X_0(t)$ . Then a parametric study was performed by varying the relevant parameters ( $T$ ,  $\dot{m}$ ,  $r_f$ , etc.) and employing the solution for the previous set of parameters as the starting vector for the succeeding set.

For transfers that required more than approximately two-thirds of a revolution, the constants  $c_1$ ,  $c_2$ ,  $c_3$ , and  $c_4$  above, were chosen to correspond to constant circumferential thrust. For shorter term transfers, these constants were chosen to correspond to an initial thrust program that is outward along the local vertical for the first half of the transit time, and inward along the local vertical for the remaining half. For a few transfers, which required many revolutions, the solution to the nonlinear state equations corresponding to constant circumferential thrust was used for the starting vector,  $X_0(t)$ . For the many revolution transfers, this choice of starting function appears more efficient than the simplified starting functions given above, even though it does not meet the boundary conditions.

For purposes of obtaining transfers that have certain particular properties it was found convenient to treat the original time optimal problem as a fixed time, maximum radius problem. This introduces a boundary condition of a more general class than previously handled within the framework of the generalized Newton-Raphson algorithm. The boundary condition appears as a nonlinear functional relation between the final value of the local horizontal velocity,  $v(t_f)$ , and the final radial distance,  $r(t_f)$ , viz:

$$\Phi(r(t_f), v(t_f)) = v^2(t_f) - k[r(t_f)]^2 = 0.$$

The procedure used for these cases was as follows: An approximate value for  $v(t_f)$  was changed automatically at each step of the iteration, on  $X_n$ , by means of the recursion formula

$$v_{n+1}(t_f) = \frac{1}{2} \left[ k r_n(t_f)^{-1} \right]^{\frac{1}{2}} \left[ 3 - \frac{r_{n+1}(t_f)}{r_n(t_f)} \right].$$

This formula results from the Newton-Raphson sequence for the scalar valued mapping  $\Phi$ , with an initial estimate  $[r_0(t_f), v_0(t_f)]$ . As  $n \rightarrow \infty$ ,  $X_n(t) \rightarrow X^*(t)$ ,  $r_n(t_f) \rightarrow r^*$ ,  $v_n(t_f) \rightarrow v^*$ , where  $X^*(t)$  is the solution of the nonlinear differential equations and  $(r^*, v^*)$  is the solution of the boundary relation  $\Phi = 0$ . This procedure was entirely systematic and exhibited good convergence properties over the range of problems studied herein.

### COMPUTATIONAL RESULTS

Computer programs utilizing the generalized Newton-Raphson method have been developed to optimize circle-to-circle transfers both for minimum time problems with specified values of final radius and for maximum radius problems with specified values of final time. The minimum time program was used to generate solutions for progressively increasing values of final radius up to durations involving 21.3 revolutions about the earth. The basic numerical data, given on the preceding pages, were used for this series of computations.

For values of final time up to a few orbital periods, the results are quite similar to those obtained from a previous near-circular linear analysis (Ref. 8). Figure 2 shows the optimum thrust steering programs for very short durations, up to one orbital period. Although the solutions shown are taken from the linear analysis of Ref. 8, the differences between these and the latest nonlinear results are at most  $2^\circ$  for the one revolution case. The time scale for each solution has been normalized so that a comparison may be made on a common scale for which the normalized time varies from zero to one. It is noted that the time variation of the thrust steering angle,  $\theta$ , is antisymmetrical with respect to the midpoint. For the very short durations, 1/6- to 1/2-revolution, the  $\theta$  motion has a mean of  $\theta = 180^\circ$  (opposite in direction to circumferential thrust), whereas the corresponding motion for durations of 2/3-revolution and longer takes place about a mean of  $\theta = 0$  (circumferential thrust). Also shown in Fig. 2 is the thrust steering angle for one revolution. For this case  $\theta$  is very nearly circumferential.

In Fig. 3 the  $\theta$  scale is considerably enlarged and a comparison is made between the linear and nonlinear solutions for the  $2\frac{1}{2}$ -revolution example. The difference is still very small, about  $\pm 3^\circ$ . However, it is noted that the duration of the last period of the nonlinear solution is slightly longer than that of the linear solution, whereas, the first periods of the two  $\theta$  time histories are almost identical in length. This characteristic is more apparent for the longer duration transfers, and is due to the thrust

FIG. 2 NORMALIZED OPTIMUM THRUST  
STEERING ANGLE FOR TRANSFER  
TIMES UP TO ONE ORBITAL PERIOD  
— LINEAR ANALYSIS —

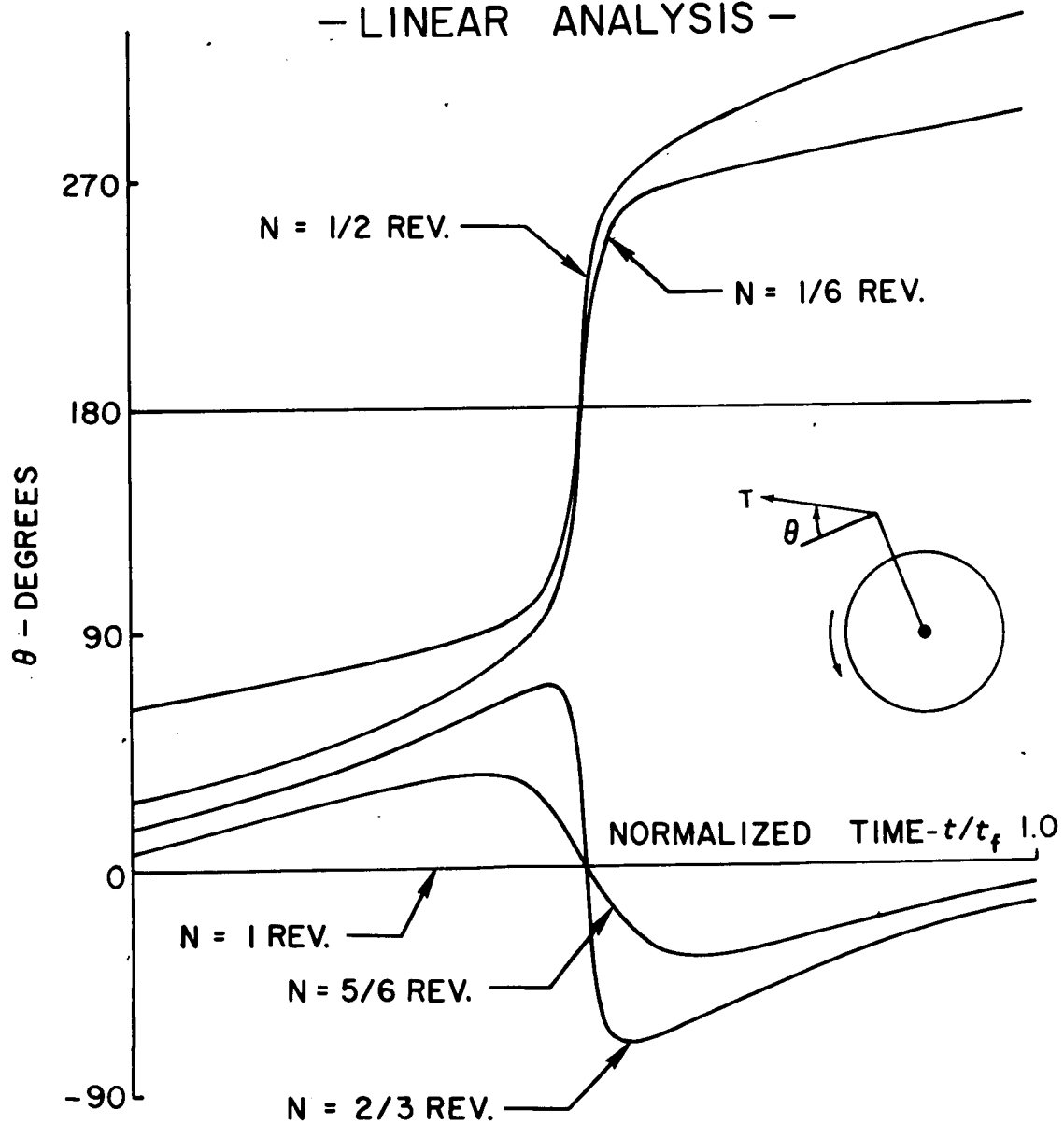
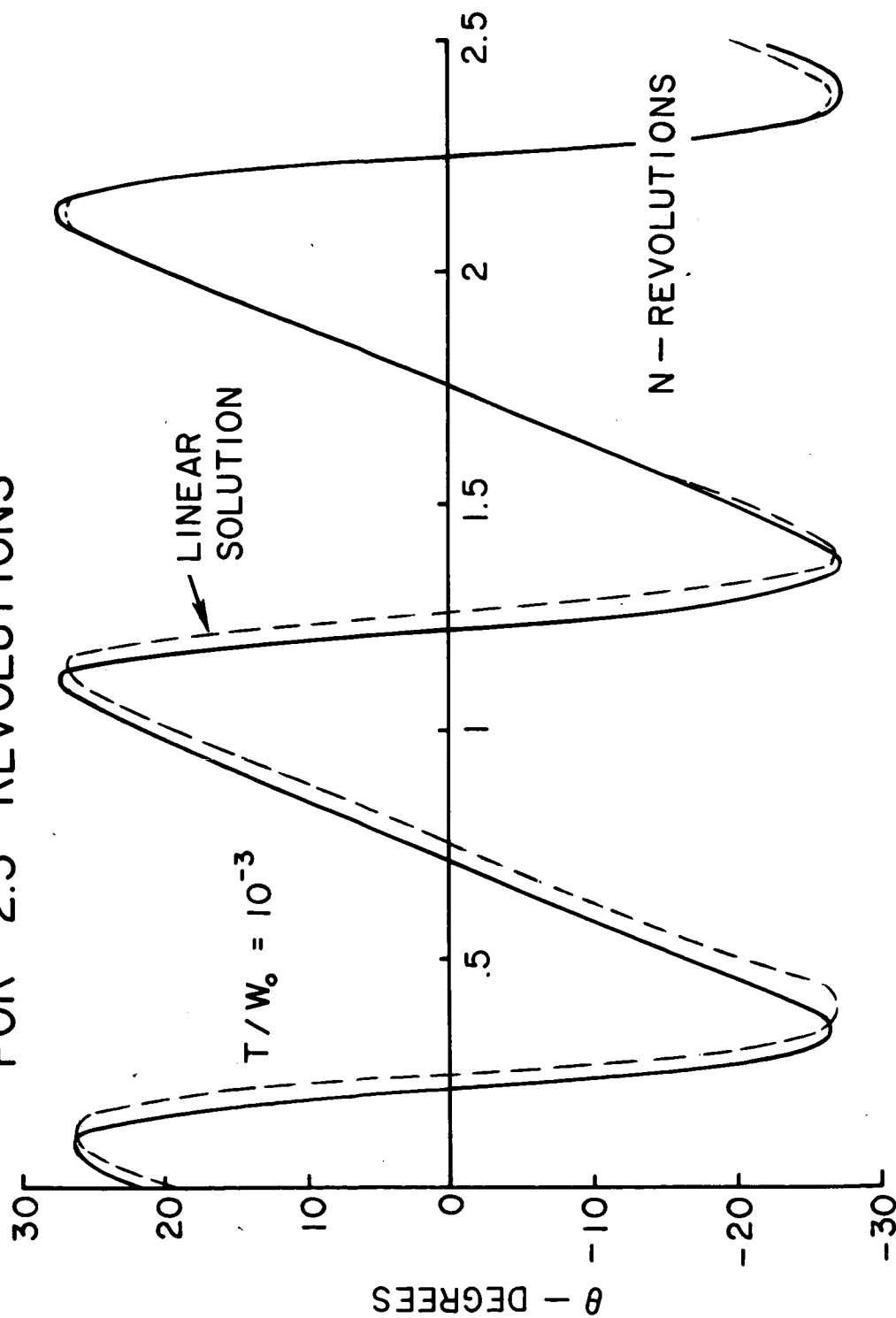


FIG. 3 OPTIMUM THRUST STEERING ANGLE  
FOR 2.5 REVOLUTIONS



steering angle  $\theta$  always being in phase with the vehicle's orbital angle,  $\varphi$  (see Fig. 1), i.e., as the altitude increases the orbital period, and therefore the period of  $\theta$  motion, also increases.

Figure 4 is also taken from the linear analysis of Ref. 8 because the differences are still relatively small. For  $N = 1\frac{1}{2}$ ,  $2\frac{1}{2}$ , and  $3\frac{1}{2}$ , the amplitude of motion is decreasing and approaching a circumferential thrust program. Also, for  $N = 1, 2$ , and  $3$ , the thrust program of the linear analysis is exactly circumferential, and only nearly circumferential for the nonlinear results of this report. A search was made for a  $\theta(t) = 0$  program for a series of solutions from  $13\frac{1}{2}$  to  $14\frac{1}{2}$  revolutions. It is clear from the results of this search that an exactly circumferential thrust program does not exist, the closest being a minimum amplitude of  $3.2^\circ$ . It has been proven independently by H. J. Kelley and R. McGill that  $\theta(t) = 0$  does not satisfy the Euler-Lagrange equations, except in the limiting case when the thrust acceleration,  $T/m$ , vanishes.

A typical optimum thrust steering program for 19 revolutions is shown in Fig. 5. As previously mentioned, the  $\theta$  motion throughout the flight is in phase with the orbital motion. Also characteristic is the relatively large  $\theta$  motion at the beginning of the transfer that diminishes to a  $3\frac{1}{2}^\circ$  to  $4\frac{1}{2}^\circ$  amplitude near the end of the maneuver. A check of the time history of eccentricity reveals that the maximum values of eccentricity build up from .0045 to .0095, and that the minimum values are very small (less than .0004) but never exactly zero except at the two termi-



FIG. 4 OPTIMUM THRUST STEERING ANGLE FOR  
1 TO 3.5 REVOLUTIONS - LINEAR ANALYSIS

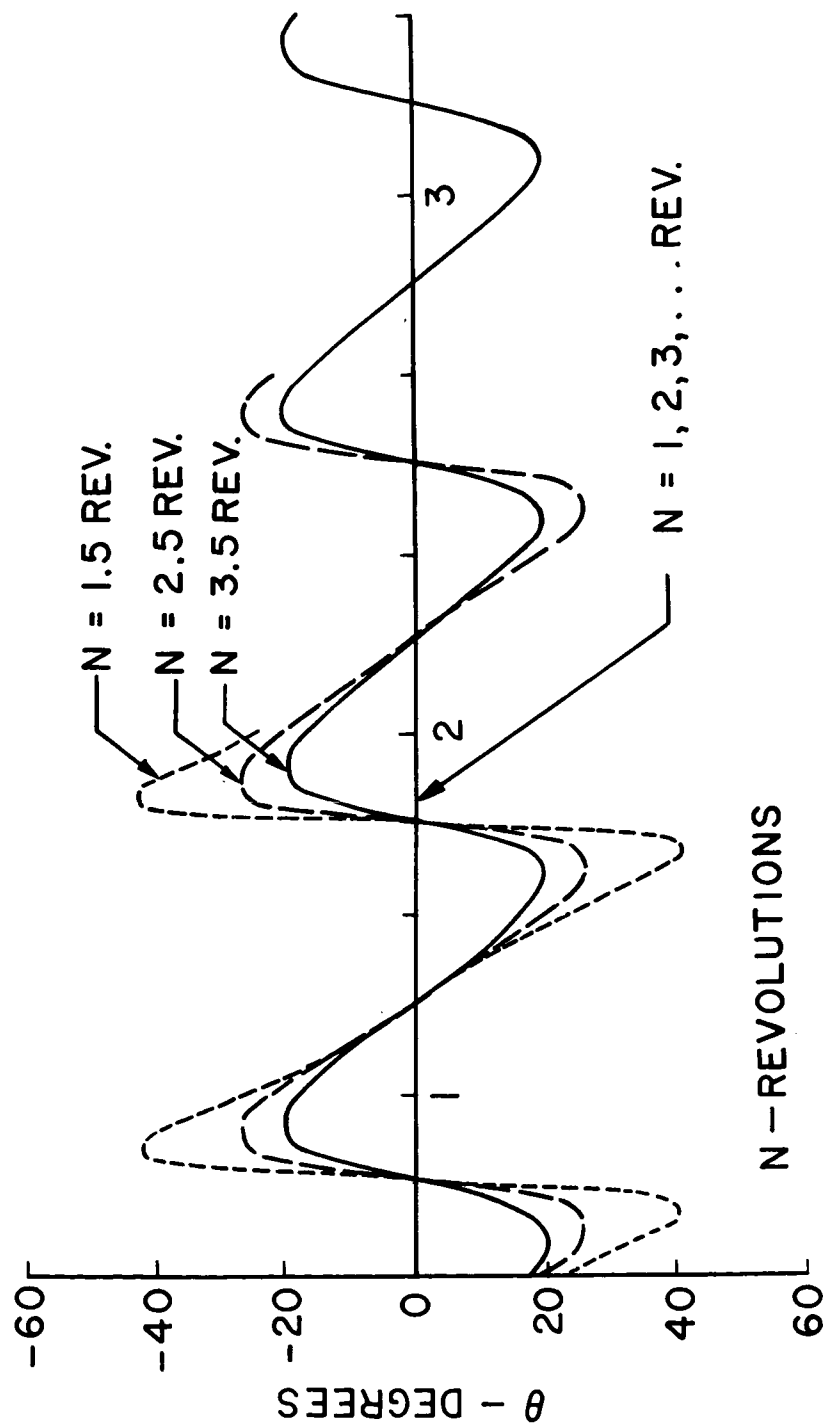
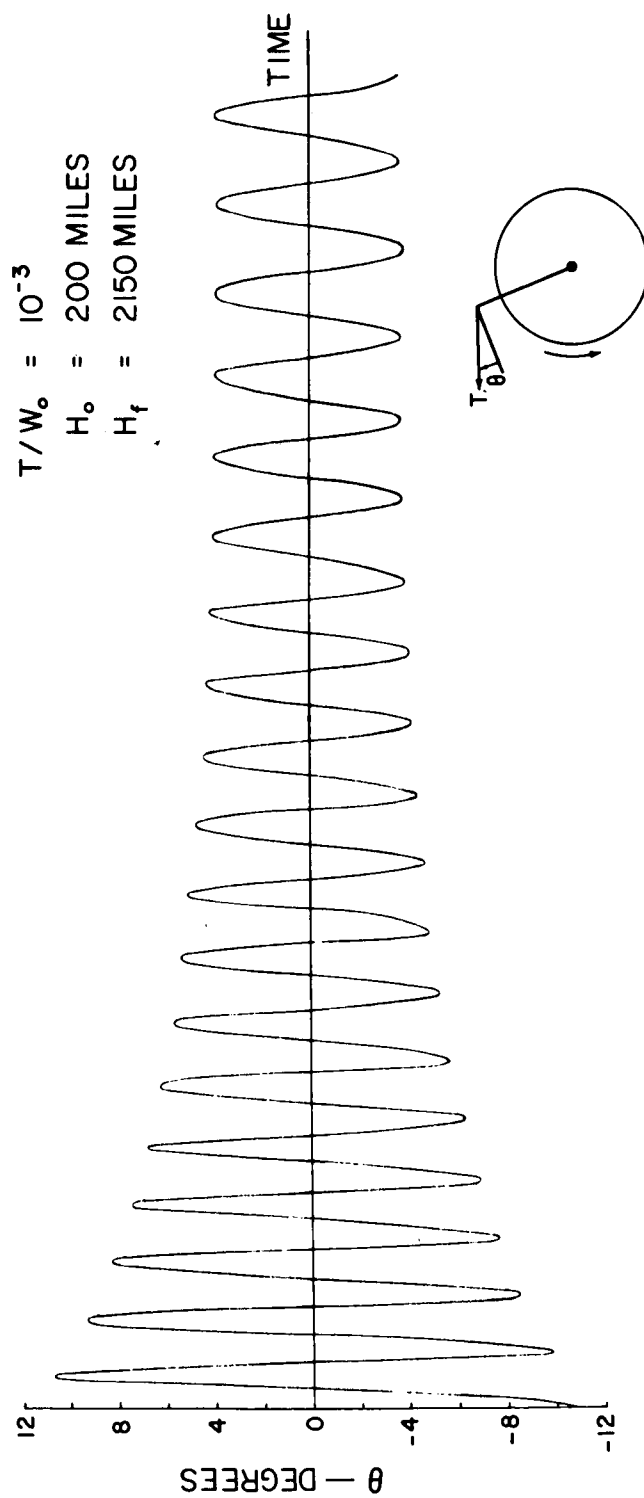


FIG 5 OPTIMUM THRUST STEERING ANGLE  
FOR 19 REVOLUTIONS



nals of the transfer. Using the generalized Newton-Raphson method, this particular solution required computation of 24 trajectories, 5 iteration cycles, 425 constant integration intervals per trajectory, an average of 22 intervals per orbit (17 intervals for the first orbit), and a total of 49 seconds of IBM 7094 computer time.

The solutions obtained are, of course, locally optimum, and no attempt has been made to search for a different class of optimum trajectories that may yield better performance. Should such a class of solutions exist, they would most likely be revealed for the significantly longer duration maneuvers.

Because the equations of motion of the linear analysis contain only the single parameter  $T/m_0 \omega_0^2$ , it is possible to plot a "miles-per-gallon" nondimensional parameter,  $\Delta r \omega_0^2 / 2\pi N_f (T/m_0)$ , as a function of the number of revolutions,  $N_f$ , required to complete the transfer. This general type of plot, taken from Ref. 8, is presented in Fig. 6. Given the thrust/mass ratio of the vehicle and the frequency of the original orbit, the increase in radius of the circular orbit may easily be computed as a function of the number of revolutions.

Similar performance results, obtained with the generalized Newton-Raphson method, are shown in Fig. 7 and do not significantly differ from the linear results. The improved performance is due to the more realistic mathematical model of the nonlinear analysis that takes into account the decrease in gravitational attraction and reduction in vehicle mass as the duration of the maneuver increases.

FIG. 6 NONDIMENSIONAL ALTITUDE GAIN PARAMETER  
- LINEAR ANALYSIS -

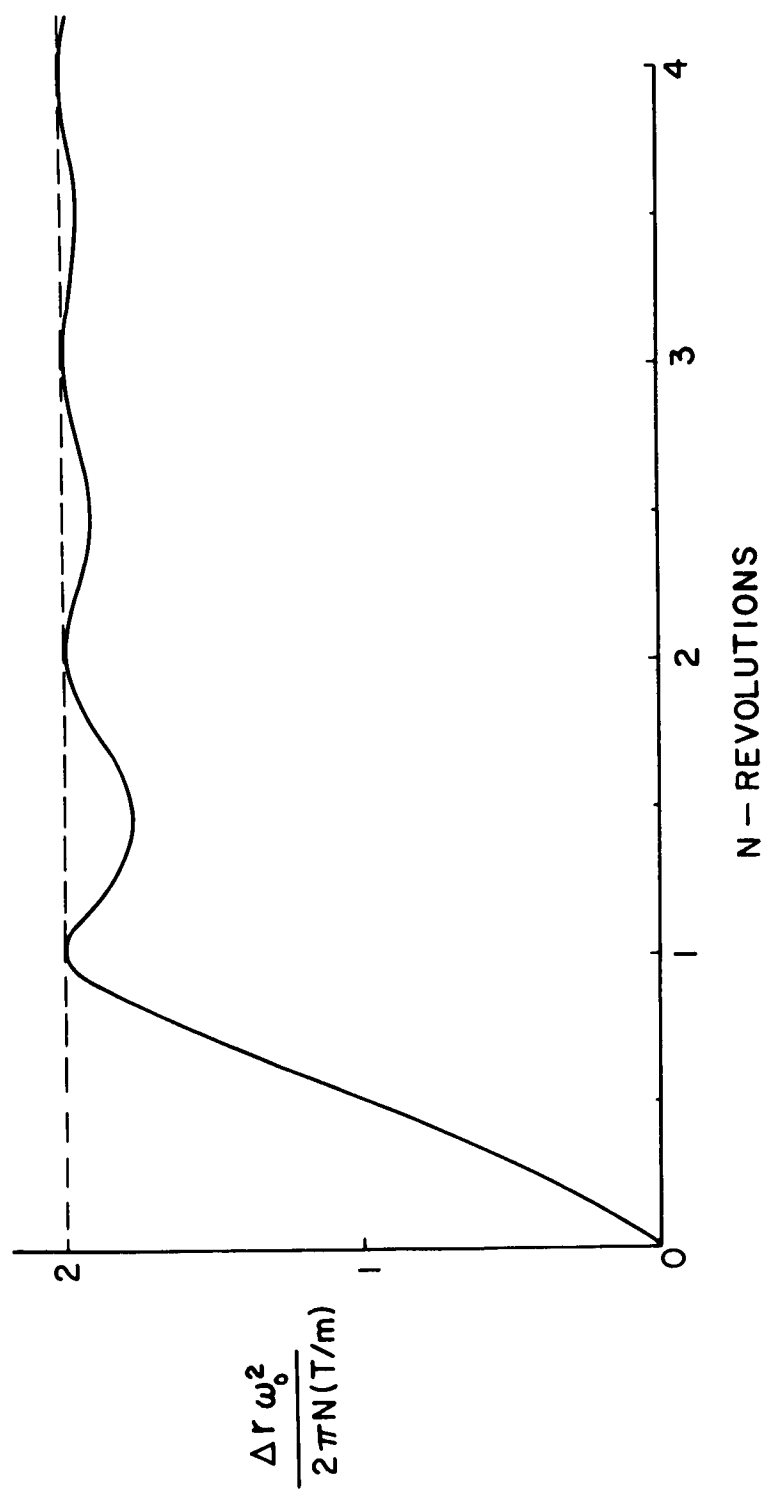
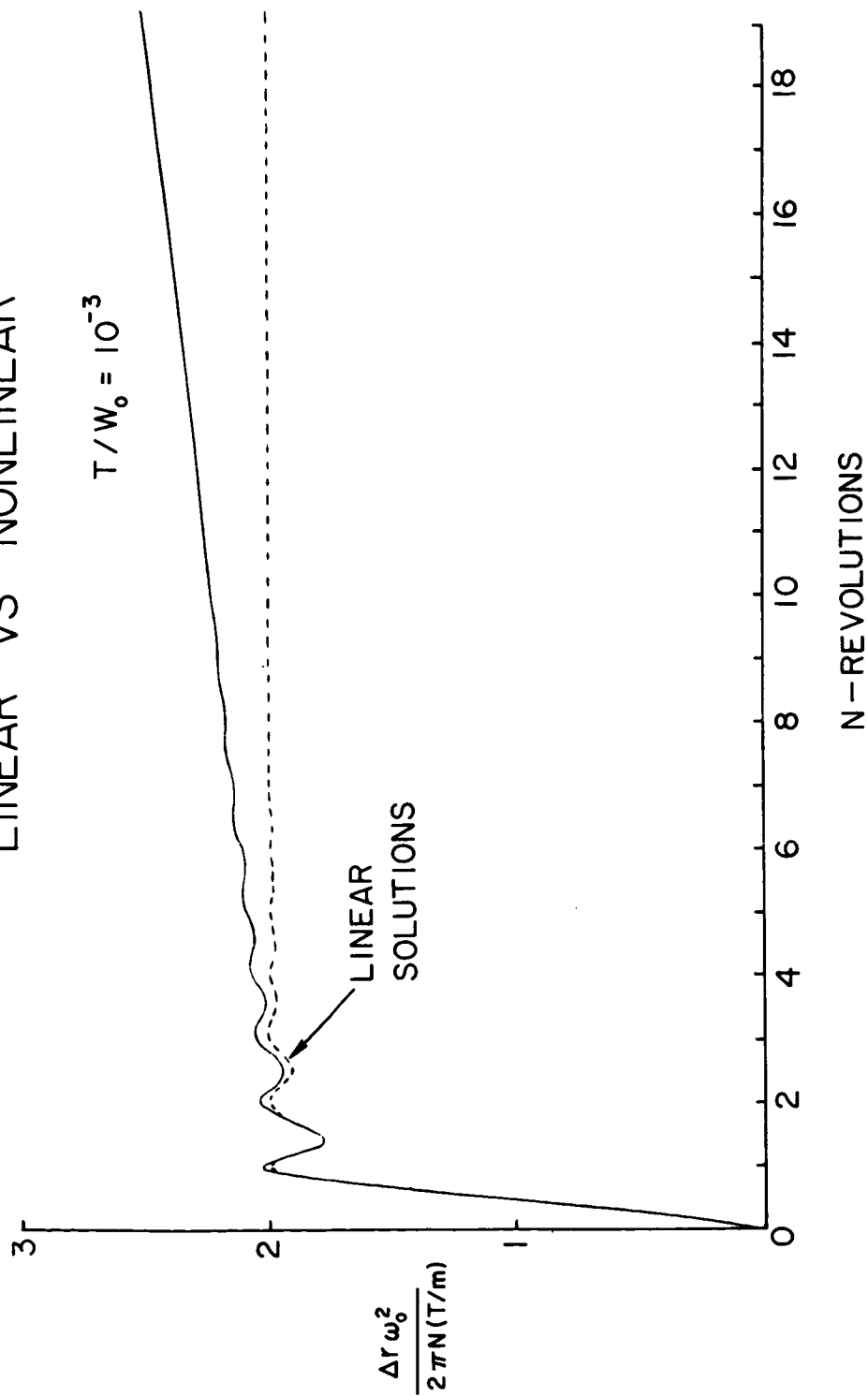


FIG. 7 NONDIMENSIONAL ALTITUDE GAIN PARAMETER  
LINEAR VS NONLINEAR



All of the previously discussed numerical results apply to a vehicle with  $T/W_0 = .001$  g's and  $I_s = 5000$  seconds. A brief vehicle parameter variation was carried out and is summarized in the following table for a fixed value of final time equal to 40.29 hours. The final time of 40.29 hours was selected because it corresponds to a transfer of 20 revolutions using the basic numerical data.

$T/W_0$	$I_s$	$\Delta R$	$N_f$
(g's)	(sec)	(miles)	(rev)
.0025	5000	10,658.	13.064
.001	5000	2,134.	20.046
.0005	5000	893.6	23.123
.00025	5000	413.9	24.792
.0001	5000	158.2	25.852
.001	1000	2,323.	19.824

Although there was no difficulty in computing an optimum transfer consisting of 21.3 revolutions, it was not possible to achieve convergence to an accuracy of four significant figures for a transfer involving 21.5 revolutions. This appears to be the limit for the generalized Newton-Raphson method employing ordinary polar coordinates and a simple second order, modified Adams, predictor-corrector, numerical integration procedure. The difficulty appears to be associated with the size and number of integration intervals rather than the number of revolutions, because there was no difficulty in computing a 26-revolution transfer with a thrust acceleration of .0001 g's.

### APPROXIMATE ANALYTICAL SOLUTIONS

In Ref. 8, optimum, low thrust transfers between neighboring circular orbits were determined for vehicles with constant thrust acceleration. It was shown that if the deviations from an original circular orbit are small, the equations may be linearized, and the resulting optimal solutions are globally minimizing. Furthermore, whenever the duration of powered flight is some integral multiple of the orbital period, the optimum thrust direction is circumferential, and the vehicle passes through a higher energy circular orbit condition at the end of each revolution.

The numerical results of the linear analysis (Ref. 8) show that for integral number of revolutions

$$\frac{y_f \omega_o^2}{\tau_f (T/m_o)} = 2 , \quad (1)$$

where  $y_f = \Delta r = r_f - r_o$  is the gain in altitude,  $\omega_o$  is the initial orbital frequency,  $\tau_f = \omega_o t_f$  is the nondimensional value of final time, and  $T/m_o$  is the constant thrust acceleration. For constant orbital frequency, the number of revolutions,  $N_f$ , is by definition

$$N_f = \frac{\omega_o t_f}{2\pi} . \quad (2)$$

Equation (1) may be rewritten as

$$\frac{\Delta r \omega_o^2}{2\pi N_f (T/m_o)} = 2 , \quad (3)$$

which is the altitude gain parameter plotted in Figs. 6 and 7.

Equation (3) may also be derived from simple energy concepts, assuming that the thrust program is circumferential. The final energy of the vehicle is expressed as the sum of the initial energy plus the work done by the rocket (work is thrust multiplied by the distance traveled,  $2\pi r_o N_f$ )

$$\frac{1}{2} m_o V_f^2 - \frac{m_o k}{r_f} = \frac{1}{2} m_o V_o^2 - \frac{m_o k}{r_o} + 2\pi r_o T N_f . \quad (4)$$

Because the initial and final orbits are circular ( $V_f^2 = k/r_f$  and  $V_o^2 = k/r_o$ ), Eq. (4) reduces to

$$\Delta r = r_f - r_o = \frac{4\pi T N_f r_o^2 r_f}{m_o k} . \quad (5)$$

Also, for neighboring circular orbits,

$$\frac{1}{\omega_o^2} = \frac{r_o^3}{k} \approx \frac{r_o^2 r_f}{k} ,$$

which reduces Eq. (5) to (3).

As a measure of performance, the following two equations, obtained from Eqs. (2) and (3), indicate the number of revolutions and time it takes for a given vehicle to transfer between circular orbits with specified radii:

$$N_f = \frac{\omega_o^2 (r_f - r_o)}{4\pi (T/m_o)} , \quad (6)$$

$$t_f = \frac{\omega_o (r_f - r_o)}{2 (T/m_o)} . \quad (7)$$



Comparison of the numerical results, based on the nonlinear mathematical model, with those obtained from Eqs. (6) and (7), shows that there is good agreement for transfers involving one or two revolutions. Thereafter, the difference between linear and nonlinear results progressively increases (see Fig. 7). For the 19-revolution example, the errors in the above linear equations are 77 per cent for  $N_f$  and 36 per cent for  $t_f$ . This is due to the assumption in the linear analysis that the gravitational attraction and mass of the vehicle are constant.

If, however,  $\omega_o$  and  $m_o$  in Eqs. (6) and (7) are continuously rectified, the new expressions should be in closer agreement with the nonlinear results. In the following derivation,  $N$  and  $t$  are considered as functions of  $r$ :

$$N = \frac{\omega_o^2 (r - r_o)}{4\pi(T/m_o)} \quad , \quad \frac{dN}{dr} = \frac{\omega_o^2}{4\pi(T/m_o)} \quad \text{for } r \approx r_o \quad ,$$

$$t = \frac{\omega_o (r - r_o)}{2(T/m_o)} \quad , \quad \frac{dt}{dr} = \frac{\omega_o}{2(T/m_o)} \quad \text{for } r \approx r_o \quad .$$

For  $r$  substantially greater than  $r_o$ , the quantity  $\omega_o^2$  is replaced by  $\omega^2 = k/r^3$ ;

$$N_f = \int_{r_o}^{r_f} \frac{dN}{dr} dr = \int_{r_o}^{r_f} \frac{(k/r^3)}{4\pi(T/m_o)} dr \quad , \quad (8)$$

$$t_f = \int_{r_0}^{r_f} \frac{dt}{dr} dr = \int_{r_0}^{r_f} \frac{(k/r^3)^{\frac{1}{2}}}{2(T/m_0)} dr, \quad (9)$$

and integration carried out with respect to  $r$ ,

$$N_f = \frac{k}{8\pi(T/m_0)} \left[ \frac{1}{r_0^2} - \frac{1}{r_f^2} \right], \quad (10)$$

$$t_f = \frac{1}{(T/m_0)} \left[ \left( \frac{k}{r_0} \right)^{\frac{1}{2}} - \left( \frac{k}{r_f} \right)^{\frac{1}{2}} \right]. \quad (11)$$

For the 19-revolution example, the errors with respect to the computed nonlinear results are reduced to 1.06 per cent for  $N_f$  and 1.23 per cent for  $t_f$ .

Because mass in the integrals of Eqs. (8) and (9) is treated as a constant, a further improvement is possible by utilizing Eq. (11) and expressing mass as a function of  $r$ :

$$m = m_0 + \dot{m}t = m_0 \left\{ 1 + \frac{\dot{m}}{T} \left[ \left( \frac{k}{r_0} \right)^{\frac{1}{2}} - \left( \frac{k}{r} \right)^{\frac{1}{2}} \right] \right\}. \quad (12)$$

If this expression is substituted in Eqs. (8) and (9) and integration is carried out again, then

$$N_f = \frac{k}{8\pi(T/m_0)} \left[ \left( 1 + \frac{\dot{m}}{T} \left\{ \frac{k}{r_0} \right\}^{\frac{1}{2}} \right) \left( \frac{1}{r_0^2} - \frac{1}{r_f^2} \right) - \frac{4}{5} \frac{\dot{m}}{T} \left( \left\{ \frac{k}{r_0^5} \right\}^{\frac{1}{2}} - \left\{ \frac{k}{r_f^5} \right\}^{\frac{1}{2}} \right) \right], \quad (13)$$

$$t_f = \frac{1}{(T/m_o)} \left[ \left( \frac{k}{r_o} \right)^{\frac{1}{2}} - \left( \frac{k}{r_f} \right)^{\frac{1}{2}} \right] \left[ 1 + \frac{\dot{m}}{2T} \left( \left\{ \frac{k}{r_o} \right\}^{\frac{1}{2}} - \left\{ \frac{k}{r_f} \right\}^{\frac{1}{2}} \right) \right] . \quad (14)$$

For the 19-revolution example, the errors are further reduced to 0.20 per cent for  $N_f$  and 0.15 per cent for  $t_f$ .

Because Eq. (14) is an improved expression for  $t$  as a function  $r$ , it is possible to repeat integration, again and again if necessary, in an attempt to reduce further the errors. This has not been carried out as the accuracy of the nonlinear results is not better than four significant figures.

## REFERENCES

1. Kelley, H.J., Hinz, H.K., Pinkham, G., and Moyer, H.G., "Low Thrust Trajectory Optimization," Progress Report No. 1 On Studies in the Fields of Space Flight and Guidance Theory, NASA-MSFC Report MTP-AERO-61-91, December 18, 1961.
2. Pinkham, G., "An Application of a Successive Approximation Scheme to Optimizing Very Low Thrust Trajectories," Progress Report No. 3 On Studies in the Fields of Space Flight and Guidance Theory, NASA-MSFC Report MTP-AERO-63-12, February 6, 1963.
3. McGill, R., and Kenneth, P., "A Convergence Theorem on the Iterative Solution of Nonlinear Two-Point Boundary Value Systems," presented at the XIV<sup>th</sup> IAF Congress, Paris, France, September 1963.
4. McGill, R., and Kenneth, P., "Solution of Variational Problems by Means of a Generalized Newton-Raphson Operator," Progress Report No. 5 On Studies in the Fields of Space Flight and Guidance Theory, NASA TM X-53024, MSFC, March 17, 1964.
5. Pontryagin, L.S., Boltyanskii, V.G., Gamkrelidze, R.V., and Mishchenko, E.F., The Mathematical Theory of Optimal Processes, Interscience Publishers, New York, 1962, p. 81.
6. Miele, A., "The Calculus of Variations in Applied Aerodynamics and Flight Mechanics," Optimization Techniques, edited by G. Leitmann, Academic Press, New York, 1962, Chapter 4.
7. Bliss, G.A., Lectures on the Calculus of Variations, University of Chicago Press, Chicago, 1946.
8. Hinz, H.K., "Optimal Low Thrust Near Circular Orbital Transfer," Progress Report No. 2 On Studies in the Fields of Space Flight and Guidance Theory, NASA-MSFC Report MTP-AERO-62-52, June 26, 1962; also AIAA Journal, p. 1367, June 1963.

RESEARCH DEPARTMENT  
GRUMMAN AIRCRAFT ENGINEERING CORPORATION

COMPUTATION OF OPTIMAL INTERPLANETARY LOW-THRUST  
TRAJECTORIES WITH BOUNDED THRUST MAGNITUDE  
BY MEANS OF THE GENERALIZED NEWTON-RAPHSON METHOD

By

Paul Kenneth  
Gerald E. Taylor

N65 33062

BETHPAGE, NEW YORK

RESEARCH DEPARTMENT  
GRUMMAN AIRCRAFT ENGINEERING CORPORATION  
BETHPAGE, NEW YORK

---

COMPUTATION OF OPTIMAL INTERPLANETARY LOW-THRUST  
TRAJECTORIES WITH BOUNDED THRUST MAGNITUDE  
BY MEANS OF THE GENERALIZED NEWTON-RAPHSON METHOD

by

Paul Kenneth  
Gerald E. Taylor

Summary

33062

The generalized Newton-Raphson method, an iterative procedure for solving nonlinear operator equations, has been extended in application to variational problems with bounded control variables. A minimum fuel interplanetary low thrust orbital transfer problem is worked out in detail to demonstrate the practical aspects of the algorithm as well as its computational effectiveness. The control variables are the thrust magnitude, limited from zero to some prescribed maximum value, and the thrust steering angle.

*Paul Taylor*

## INTRODUCTION

For variational problems, not involving inequality constraints on state or control variables, the state equations and Euler-Lagrange equations generally consist of a system of nonlinear differential equations with two-point boundary conditions. For such a system, the generalized Newton-Raphson technique proceeds by solving a sequence of linear boundary value problems in such a manner that the sequence of solutions converges to the solution of the nonlinear boundary value problem. The generalized Newton-Raphson operator technique has been developed for such systems of ordinary differential equations with two-point boundary conditions (Ref. 1) and successfully applied to various unconstrained variational problems (Ref. 2).

In this paper, we consider variational problems with inequality constraints on at least one control variable. Following Valentine (Ref. 3), a new variable is introduced such that the inequality constraint may be replaced by an equivalent equality constraint. The resulting nonlinear system of state and Euler-Lagrange equations now consists of differential equations and algebraic equations. The generalized Newton-Raphson method is applied to this nonlinear

operator equation. Again, this is accomplished by solving a sequence of linear operator equations such that the sequence of solutions converges to the solution of the nonlinear operator equation.

The algorithm is applied to the computation of minimum fuel, low-thrust, Earth to Mars orbit transfer trajectories, with bounded thrust magnitude.



## PROBLEM FORMULATION

Given the differential constraints

$$\varphi_i = \dot{x}_i - f_i(t, x_1, \dots, x_n, u_1, \dots, u_m) = 0, \quad (1)$$

$$i = 1, 2, \dots, n,$$

and at most  $2n+1$  end conditions involving  $t$  and  $x_i$ , as well as inequality constraints

$$u_{K_{\min}} \leq u_K \leq u_{K_{\max}}, \quad K = 1, \dots, r \leq m,$$

the problem is to determine the state variables  $x_i(t)$  and control variables  $u_j(t)$  so as to minimize the function

$$P = P(t_0, t_f, x_1(t_0), \dots, x_n(t_0), x_1(t_f), \dots, x_n(t_f)).$$

A set of new real variables  $\alpha_K$  is introduced (Refs. 3-6), and the inequality constraints on the control variables are replaced by

$$\Phi_K = (u_K - u_{K_{\min}})(u_{K_{\max}} - u_K) - \alpha_K^2 = 0 \quad (2)$$

$$K = 1, \dots, r \leq m.$$

Let

$$F \equiv \sum_{i=1}^n \lambda_i(t) \varphi_i + \sum_{K=1}^r \lambda_K(t) \Phi_K = 0, \quad (3)$$

where the  $\lambda(t)$  are undetermined multipliers. From a modification of the classical calculus of variations (Refs. 4-8), we obtain as necessary conditions for the existence of a local minimum of  $P$ :

(a) the Euler-Lagrange equations

$$\begin{aligned}\frac{d}{dt} \frac{\partial F}{\partial \dot{x}_i} - \frac{\partial F}{\partial x_i} &= 0, \quad i = 1, 2, \dots, n \\ \frac{\partial F}{\partial u_j} &= 0, \quad j = 1, 2, \dots, m \\ \frac{\partial F}{\partial \alpha_K} &= 0, \quad K = 1, 2, \dots, r,\end{aligned}\tag{4}$$

(b) the transversality conditions

$$dP + \left[ \sum_{i=1}^n \frac{\partial F}{\partial \dot{x}_i} dx_i + \left( F - \sum_{i=1}^n \dot{x}_i \frac{\partial F}{\partial \dot{x}_i} \right) dt \right]_{t_0}^{t_f} = 0,$$

where the  $dt$  and  $dx_i$  are differentials which are connected by the prescribed end conditions,

(c) the Weierstrass condition, which for this problem is equivalent (Ref. 4) to the requirement that

$$H \equiv \sum_{i=1}^n \lambda_i f_i(t, x_1, \dots, x_n, u_1, \dots, u_m)$$

be maximum with respect to the control variables  $u_j$  satisfying the imposed inequality constraints.

To obtain the solution of the problem stated above, the generalized Newton-Raphson algorithm is applied to the operator equation consisting of Eqs. (1), (2), and (4). This operator equation consists of a two-point boundary value system of order  $2n$ , in addition to a system of scalar equations of order  $m + 2r$ . The following numerical example should clarify the computational procedure.

#### LOW-THRUST ORBITAL TRANSFER EXAMPLE — MINIMUM FUEL

The problem we wish to consider is closely related to the last example in Ref. 2. Kelley et al. (Ref. 9) have obtained results to the minimum time version of the problem via gradient techniques. We wish to minimize the fuel consumption of a low-thrust ion rocket which is to transfer from the orbit of Earth to the orbit of Mars, in fixed time. The orbits of Earth and Mars are assumed to be circular and coplanar, and the gravitational attractions of the two planets are neglected. The system parameters are: the initial mass  $m_0$ , 46.58 slugs; the constant equivalent exit velocity,  $c = 1.831 \times 10^5$  ft/sec; and the propellant mass flow,  $\beta$ , which is required to remain within the bounds  $\beta_{\max} = 6.937 \times 10^{-7}$  slugs/sec, and  $\beta_{\min} = 0$ .

We now proceed to the formal statement of the problem.  
Given the differential constraints:

$$\begin{aligned}
 \dot{r} &= w \\
 \dot{w} &= \frac{v^2}{r} - \frac{K}{r^2} + \frac{c\beta}{m} \sin \theta \\
 \dot{v} &= -\frac{wv}{r} + \frac{c\beta}{m} \cos \theta \\
 \dot{m} &= -\beta
 \end{aligned}
 \tag{5}$$

with the boundary conditions:

$$\begin{aligned}
 t &= t_0 & t &= t_f \\
 r(t_0) &= r_0 & r(t_f) &= r_f \\
 w(t_0) &= w_0 & w(t_f) &= w_f \\
 v(t_0) &= v_0 & v(t_f) &= v_f \\
 m(t_0) &= m_0 & m(t_f) &\sim \text{open} ,
 \end{aligned}$$

where  $w$  and  $v$  are the radial and circumferential velocities respectively;  $r$  is the radius; and  $\theta$  is the thrust direction angle measured from the local horizontal. In addition, given the inequality constraints

$$\beta_{\min} \leq \beta \leq \beta_{\max} ,$$

determine the state variables  $r(t)$ ,  $w(t)$ ,  $v(t)$ ,  $m(t)$  and control variables  $\theta(t)$  and  $\beta(t)$  so as to minimize

$$P = - m(t_f) .$$

Rewriting the inequality constraints on  $\beta$  as

$$(\beta - \beta_{\min})(\beta_{\max} - \beta) - \alpha^2 = 0 ,$$

the Euler-Lagrange equations, Eqs. (4), become

$$\begin{aligned} \dot{\lambda}_r &= \lambda_w \left( \frac{v^2}{r^2} - \frac{2K}{r^3} \right) - \lambda_v \frac{wv}{r^2} \\ \dot{\lambda}_w &= \lambda_v \frac{v}{r} - \lambda_r \\ \dot{\lambda}_v &= - 2\lambda_w \frac{v}{r} + \lambda_v \frac{w}{r} \\ \dot{\lambda}_m &= \frac{c\beta}{m^2} (\lambda_w \sin \theta + \lambda_v \cos \theta) \\ 0 &= \frac{c\beta}{m} (-\lambda_w \cos \theta + \lambda_v \sin \theta) \\ 0 &= \frac{c}{m} (\lambda_w \sin \theta + \lambda_v \cos \theta) - \lambda_m - \lambda_\alpha (\beta_{\max} + \beta_{\min} - 2\beta) \\ 0 &= \alpha \lambda_\alpha . \end{aligned} \tag{6}$$

Equations (6) and the Weierstrass condition imply

$$\begin{aligned} \sin \theta &= \lambda_w \left( \lambda_w^2 + \lambda_v^2 \right)^{-\frac{1}{2}} \\ \cos \theta &= \lambda_v \left( \lambda_w^2 + \lambda_v^2 \right)^{-\frac{1}{2}} . \end{aligned} \tag{7}$$

Substitution of Eqs. (7) into Eqs. (5) and Eqs. (6), now yields the nonlinear boundary value problem

$$\begin{aligned}
 \dot{r} &= w & &= f^1 \\
 \dot{w} &= \frac{v^2}{r} - \frac{K}{r^2} + \frac{c\beta}{m} \lambda_w \left( \lambda_w^2 + \lambda_v^2 \right)^{-\frac{1}{2}} & &= f^2 \\
 \dot{v} &= -\frac{wv}{r} + \frac{c\beta}{m} \lambda_v \left( \lambda_w^2 + \lambda_v^2 \right)^{-\frac{1}{2}} & &= f^3 \\
 \dot{m} &= -\beta & &= f^4 \\
 \dot{\lambda}_r &= \lambda_w \left( \frac{v^2}{r^2} - \frac{2K}{r^3} \right) - \lambda_v \frac{wv}{r^2} & &= f^5 \\
 \dot{\lambda}_w &= \lambda_v \frac{v}{r} - \lambda_r & &= f^6 \\
 \dot{\lambda}_v &= -2\lambda_w \frac{v}{r} + \lambda_v \frac{w}{r} & &= f^7 \\
 \dot{\lambda}_m &= \frac{c\beta}{m^2} \left( \lambda_w^2 + \lambda_v^2 \right)^{\frac{1}{2}} & &= f^8,
 \end{aligned} \tag{8}$$

with boundary conditions

$$\begin{aligned}
 t &= t_0 & t &= t_f \\
 r(t_0) &= r_0 & r(t_f) &= r_f \\
 w(t_0) &= w_0 & w(t_f) &= w_f \\
 v(t_0) &= v_0 & v(t_f) &= v_f \\
 m(t_0) &= m_0 \\
 \lambda_r(t_0) &= \lambda_{r_0} \text{ The constant } \lambda_{r_0} \text{ scales the multipliers.}
 \end{aligned}$$

In addition to the boundary value system given by Eqs. (8), we have to satisfy the equations

$$\begin{aligned}
 0 &= (\beta - \beta_{\min})(\beta_{\max} - \beta) - \alpha^2 & &= g^9 \\
 0 &= \frac{c}{m} \left( \lambda_w^2 + \lambda_v^2 \right)^{\frac{1}{2}} - \lambda_m - \lambda_\alpha (\beta_{\max} + \beta_{\min} - 2\beta) & &= g^{10} \\
 0 &= \alpha \lambda_\alpha & &= g^{11} .
 \end{aligned} \tag{9}$$

For the discussion of the application of the Newton-Raphson operator technique to the nonlinear system consisting of Eqs. (8) and Eqs. (9), we rewrite these equations as follows:

$$\begin{aligned}
 \dot{X} &= F(X, t) \quad , \quad t \in [t_0, t_f] \\
 G(X, t) &= 0 \quad ,
 \end{aligned} \tag{10}$$

where

$$\begin{aligned}
 X &= (x^1, \dots, x^{11}) \\
 F &= (f^1, \dots, f^8) \\
 G &= (g^9, g^{10}, g^{11}) \\
 f^i &= f^i(x^1, \dots, x^{11}, t) \quad , \quad i = 1, \dots, 8 \\
 g^i &= g^i(x^1, \dots, x^{11}, t) \quad , \quad i = 9, 10, 11 ,
 \end{aligned}$$

and

$$\begin{aligned}
x^1(t) &= r(t) \quad , \quad x^2(t) = w(t) \quad , \quad x^3(t) = v(t) \quad , \\
x^4(t) &= m(t) \quad , \quad x^5(t) = \lambda_r(t) \quad , \quad x^6(t) = \lambda_w(t) \quad , \\
x^7(t) &= \lambda_v(t) \quad , \quad x^8(t) = \lambda_m(t) \quad , \quad x^9(t) = \beta(t) \quad , \\
x^{10}(t) &= \alpha(t) \quad , \quad x^{11}(t) = \lambda_\alpha(t) \quad .
\end{aligned}$$

The algorithm now requires the solution of the sequence of linear equations:

$$\dot{X}_{n+1} = J(X_n, t) (X_{n+1} - X_n) + F(X_n, t) \quad (11a)$$

$$0 = I(X_n, t) (X_{n+1} - X_n) + G(X_n, t) \quad (11b)$$

$$n = 0, 1, \dots,$$

where  $J(X, t)$  is the matrix with elements  $J_{ij} = \frac{\partial f^i}{\partial x^j}$ ,  
 $i = 1, \dots, 8, j = 1, \dots, 11$ ; and  $I(X, t)$  is the matrix  
with elements  $I_{ij} = \frac{\partial g^i}{\partial x^j}$ ,  $i = 9, 10, 11, j = 1, \dots, 11$ .

At every iterate  $n$ ,  $x_n^9 = x_n^9(x_n^1, \dots, x_n^8)$  is obtained from Eq. (11b). This relation is used to eliminate  $x_n^9$  from Eq. (11a). The functions  $x_n^1(t), \dots, x_n^8(t)$  are then computed from Eq. (11a), after which  $x_n^9(t), x_n^{10}(t), x_n^{11}(t)$  are computed from Eq. (11b). A description of the



method of solution for the linear two-point boundary value system, Eq. (11a), with the given end conditions is contained in Ref. 2. The iteration proceeds until  $\rho(X_{n+1}, X_n) \leq \delta$ , where

$$\rho(X_{n+1}, X_n) = \sum_{i=1}^8 \max_{t \in [t_0, t_f]} |x_{n+1}^i(t) - x_n^i(t)|, \quad (12)$$

and  $\delta$  is a suitably small positive constant. The corresponding iterate  $X_{n+1}$  is accepted as a solution, and a final check is made by integrating the nonlinear Eqs. (8) with a complete set of initial conditions taken from the final iterate, and with  $\beta(t)$  computed at every integration step by

$$\beta = \begin{cases} \beta_{\max} & , \quad \text{when } \eta \geq 0 \\ \beta_{\min} & , \quad \text{when } \eta < 0 \end{cases}, \quad (13)$$

where

$$\eta = \frac{c}{m} \left( \lambda_w^2 + \lambda_v^2 \right)^{\frac{1}{2}} - \lambda_m.$$

Equation (13) results from the Weierstrass condition, viz., maximizing  $H$  with respect to  $\beta$ .

The data for the problem are normalized to obtain:

$$\begin{array}{ll}
 r_0 = 1.000 & r_f = 1.525 \\
 w_0 = 0.000 & w_f = 0.000 \\
 v_0 = 1.000 & v_f = 0.8098 \\
 m_0 = 1.000 & \beta_{\max} = 0.07500 \\
 \lambda_{r_0} = 1.000 & \beta_{\min} = 0.000 \\
 K = 1.000 & c = 1.872 .
 \end{array}$$

This results in a time unit of 58.18 days. The final time  $t_f$  is chosen to be 3.816 units (222.0 days). The starting vector  $X_0(t)$  is chosen as follows:

$$\begin{aligned}
 x_0^1(t) &\equiv r_0(t) = r_0 + \frac{r_f - r_0}{t_f} t \\
 x_0^2(t) &\equiv w_0(t) \equiv 0 \\
 x_0^3(t) &\equiv v_0(t) = \left[ \frac{K}{r_0(t)} \right]^{\frac{1}{2}} \\
 x_0^4(t) &\equiv m_0(t) = 1 - \frac{t}{4t_f} \\
 x_0^5(t) &\equiv \lambda_{r_0}(t) \equiv 1.000
 \end{aligned} \tag{14}$$

$$x_0^6(t) \equiv \lambda_{w_0}(t) = \begin{cases} 0.5200 & , \quad \text{for } t \in [0, \frac{1}{2}t_f] \\ -0.5000 & , \quad \text{for } t \in (\frac{1}{2}t_f, t_f] \end{cases}$$

$$x_0^7(t) \equiv \lambda_{v_0}(t) = \begin{cases} 0.3000 & , \quad \text{for } t \in [0, \frac{1}{2}t_f] \\ 0.000 & , \quad \text{for } t \in (\frac{1}{2}t_f, t_f] \end{cases}$$

$$x_0^8(t) \equiv \lambda_{m_0}(t) \equiv 0 \quad (14)$$

(Cont.)

$$x_0^9(t) \equiv \beta_0(t) = \frac{\beta_{\max}}{2} \left( 1 + \cos \frac{2\pi t}{t_f} \right)$$

$$x_0^{10}(t) \equiv \alpha_0(t) = \frac{\beta_{\max}}{10} \left( 1 + \frac{t}{t_f} \right)$$

$$x_0^{11}(t) \equiv \lambda_{\alpha_0}(t) = 10 \sin \frac{\pi t}{t_f} - 11 .$$

To carry out the necessary computations, the time interval  $[t_0, t_f]$  is divided into 200 equal subintervals. After the switching times ( $\eta = 0$ ) have been located, within the accuracy of the grid size, the time steps in the neighborhood of the switching points are further subdivided into 10 equal intervals, and the iterations continued with this refined grid. In this manner, it is possible to locate the switching points, or points of discontinuity of the control  $\beta(t)$ , with greater precision.

The sequence  $\{X_n\}$  converges to an accuracy of 4 significant figures in 51 total iterations. The total computer time (IBM 7094) required is approximately 2 minutes. Figures 1 and 2 illustrate the convergence for the control variables  $\theta(t)$  and  $\beta(t)$  respectively.  $\theta^*(t)$  and  $\beta^*(t)$  result from the final check of the nonlinear state and Euler-Lagrange equations, Eqs. (8), with the switching points obtained from Eq. (13).

With the same starting vector  $X_0(t)$ , Eqs. (14), trajectories have also been computed with final times  $t_f = 195.0, 201.0, 208.0, 215.0$  days. In Fig. 3 the final time  $t_f$  is plotted against the ratio of final mass  $m_f$  to initial mass  $m_0$ .

#### CONCLUDING REMARKS

With the integration routine utilized for these sample problems, the solutions seem to be limited to an accuracy of 4 significant figures. We believe that through the use of higher precision integration schemes, presently under investigation at Grumman, more accurate results can be obtained.

### ACKNOWLEDGMENT

The authors wish to thank Dr. Henry J. Kelley for his crucial suggestions pertaining to this work.

Portions of the theoretical part of this study were generated in connection with AFOSR Contract No. AF49(638)-1207.

### REFERENCES

1. McGill, R., and Kenneth, P., "A Convergence Theorem on the Iterative Solution of Nonlinear Two-Point Boundary Value Systems," presented at the XIV<sup>th</sup> IAF Congress, Paris, France, September 1963.
2. McGill, R., and Kenneth, P., "Solution of Variational Problems by Means of a Generalized Newton-Raphson Operator," in Progress Report No. 5 on Studies in the Fields of Space Flight and Guidance Theory, MSFC, Huntsville, Alabama, March 1964.
3. Valentine, F.A., "The Problem of Lagrange with Differential Inequalities as Added Side Conditions," Dissertation, Department of Mathematics, University of Chicago, Chicago, Illinois, 1937.
4. Leitmann, G., "Variational Problems with Bounded Control Variables," Chapter 5 of Optimization Techniques, edited by G. Leitmann, Academic Press, New York, 1962.
5. Leitmann, G., "An Elementary Derivation of the Optimal Control Conditions," Proceedings of the XII<sup>th</sup> IAF Congress, Washington, D.C., 1961.
6. Berkovitz, L.D., "Variational Methods in Problems of Control and Programming," J. Math. Anal. and Appl. 3, 145 (1961).

7. Bliss, G.A., Lectures on the Calculus of Variations, University of Chicago Press, Chicago, 1946.
8. Breakwell, J.V., "The Optimization of Trajectories," J. Soc. Ind. Appl. Math. 7, 215 (1959).
9. Kelley, H.J., Kopp, R.E., and Moyer, H.G., "Successive Approximation Techniques for Trajectory Optimization," presented at the IAS Vehicle Systems Optimization Symposium, Garden City, New York, November 1961.

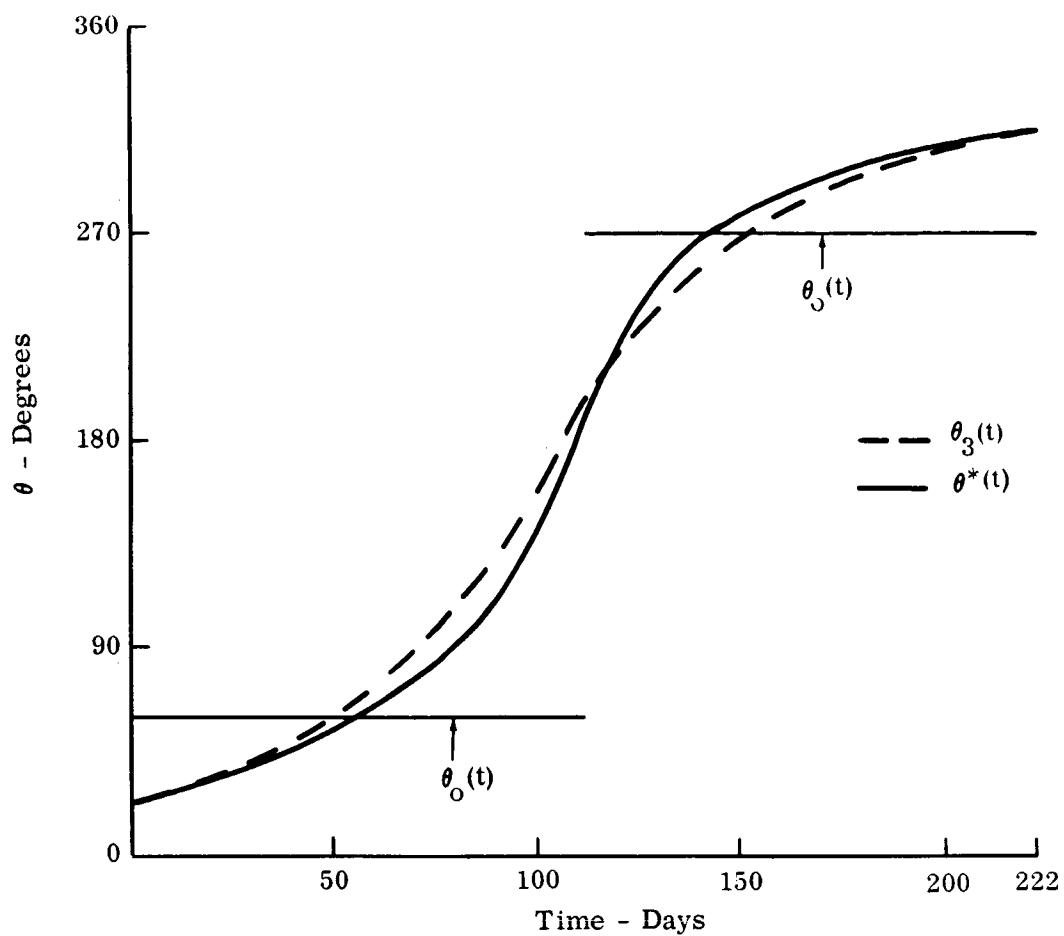


Figure. 1 - Control Angle Histories

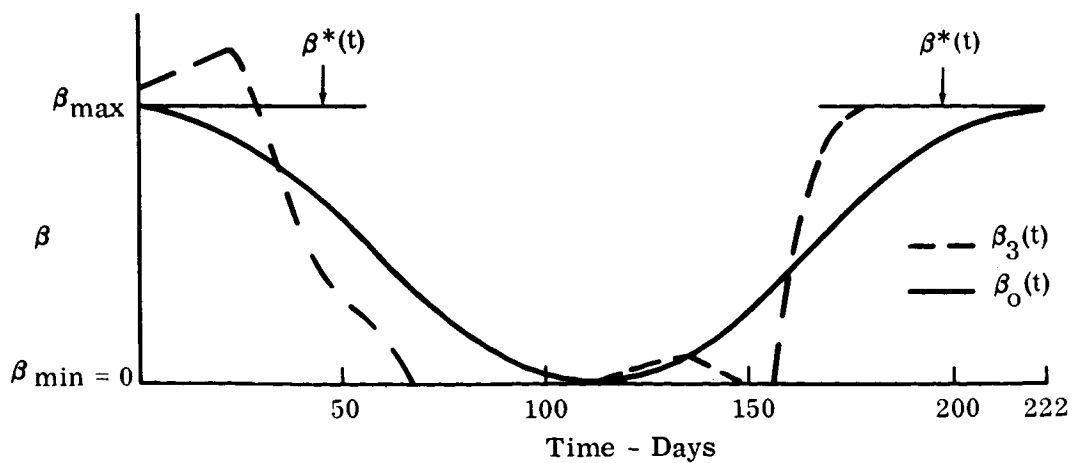


Fig. 2 - Throttle Control Histories

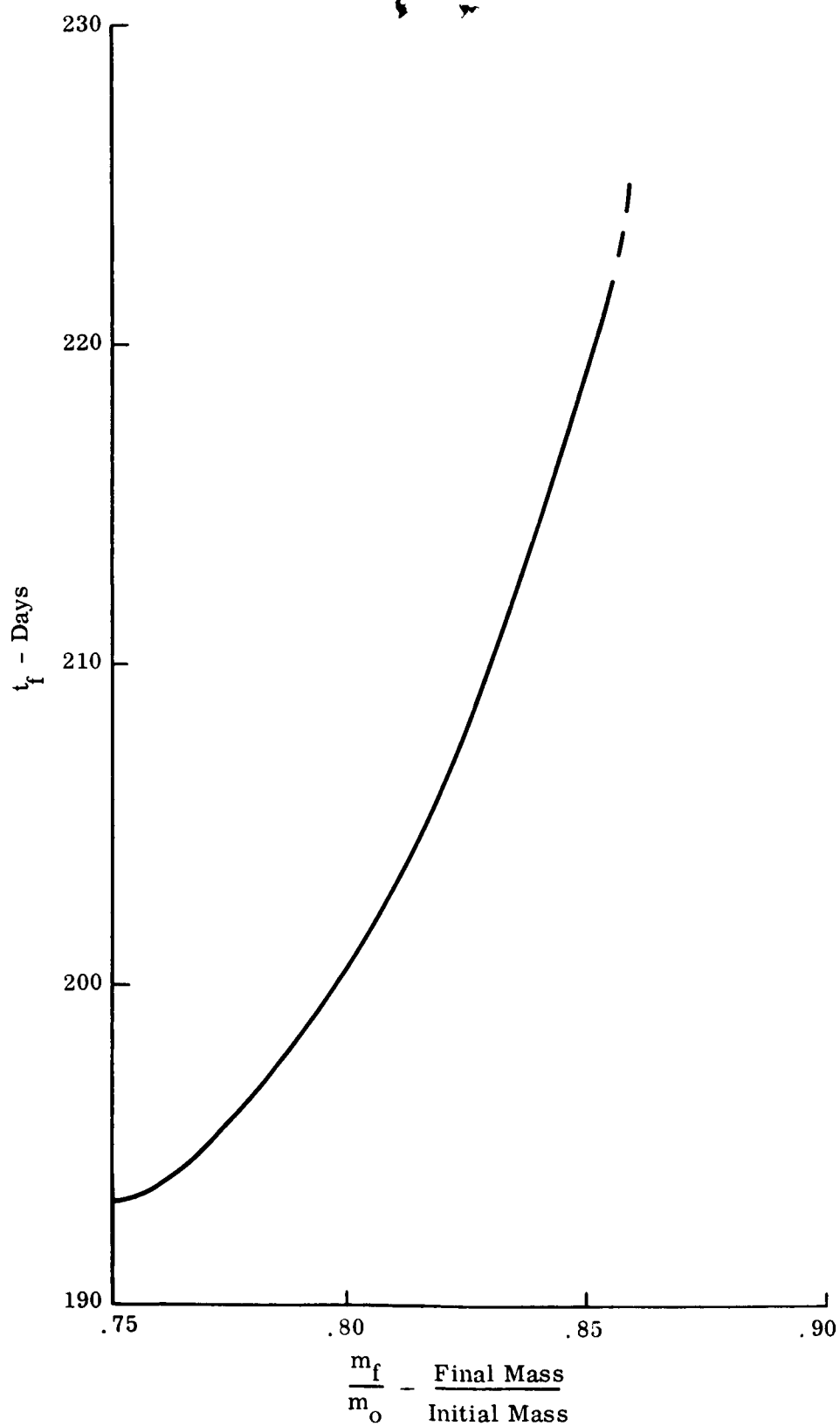


Fig. 3 - Transfer Time versus Final Mass



APPROXIMATING OPTIMAL TRAJECTORIES:  
SELECTION OF SIGNIFICANT ESTIMATION VARIABLES  
IN A LEAST SQUARES PROBLEM

I. E. PERLIN, J. H. MACKAY,  
J. W. WALKER, J. J. GOODE, O. B. FRANCIS, JR.

RICH COMPUTER CENTER.  
GEORGIA INSTITUTE OF TECHNOLOGY  
ATLANTA, GEORGIA

N65 33063

ABSTRACT

33063

In this report a step-up procedure for the selection of significant estimation variables in a least squares problem is developed. Application of this procedure to several examples is made, and a computer program in ALGOL 58 compiler language for the Burroughs 220 computer is discussed.

Author

APPROXIMATING OPTIMAL TRAJECTORIES: SELECTION OF SIGNIFICANT  
ESTIMATION VARIABLES IN A LEAST SQUARES PROBLEM

The Astrodynamics and Guidance Theory Division of the Aero-Astrodynamics Laboratory of the Marshall Space Flight Center is examining the role of "large computers" as they may be exploited in the control and guidance of missile performance. Under Contract No. NAS8-5365 the Georgia Institute of Technology and its Rich Electronic Computer Center have been studying such exploitation as it applies to the approximation of guidance functions with multivariate functional models. Under this contract attention so far has been focused on methods to reduce the computational and variable-selection problems in least squares models.

Background

The state vector,  $x(t)$  (describing the flight of a missile through space) has the derivative  $\dot{x}(t)$ . These vectors along with a vector descriptive of the guidance function,  $u(t)$ , satisfy equations of motion, which may be expressed formally as

$$F[\dot{x}(t), x(t), t, u(t)] = 0$$

The missile is intended to satisfy certain mission requirements at some future time,  $t_c$ , and we may indicate these requirements in the equations describing terminal conditions:

$$G[x(t_c), \dot{x}(t_c), t_c] = 0$$

Note that the functions  $F$  and  $G$  are themselves vectors. The guidance problem may be expressed generally as that of choosing a "best" guidance function  $u$  out of the class of possible guidance functions. In particular we may wish to choose a function  $u$  in such a way as to minimize

$$\int_0^{t_c} c(x, \dot{x}, u, t) dt$$

In practical situations with real missiles we could not use the exact optimum guidance function as a function of time because of measurement errors and so on. The missile strays from the optimum path into a situation for which the chosen guidance function is no longer best. It then becomes necessary to calculate a new optimum guidance function based on new initial conditions. In short it is important to be able to synthesize the optimal guidance function,  $u$ , in terms of the state variables at each point in the phase space.

One approach to this synthesis which has been proposed consists in selecting a scatter of initial points (possibly organized in subregions of the phase space); using a large-scale computer to determine the corresponding values of the optimal guidance function; and then using some approximation technique to estimate the guidance function as a function of the state of the missile.

Various considerations, both practical and theoretical, suggest that such an approximation be based on the criterion of "least squares." Even, however, if attention is restricted to this well-known method, difficulties arise. In the first place fitting a function of several variables becomes very quickly a huge matrix inversion problem. In an earlier study done under this contract, entitled: "Least Squares Estimation of Regression Coefficients in a Special Class of Polynomial Models," techniques were described which reduced the large inversion problem to a sequence of low-order inversions, when fitting balanced polynomials to rectangular grids of data. While these techniques hold promise in special circumstances, evidently they have a limited usefulness.

A second major difficulty in least squares approximations arises in deciding which class of functions or which subset of a very large class of estimation variables will be used to approximate the unknown function. Evidently, a method which elects a relatively few highly efficient estimation variables also serves to keep the matrix-inversion problem under control, since that computation depends directly on the number of estimation variables used.

It happens that there is a method available by means of which the incorporation of estimation variables into the approximating functions can be sequenced in what seems usually to be an efficient manner. We shall call this formal procedure for activating estimation variables simply the step-up procedure. The procedure appears first to have been used by R. J. Wherry (Annal. of Math. Stat., 1931). More recent discussions have appeared by H. E. Anderson and B. Fruchter (Psychometrika, 1960), and E. F. Schultz, Jr. and J. F. Goggans (Bulletin of the Agricultural Exp. Station, Auburn Univ., 1961). Since examples can be constructed to show that the step-up procedure is not always optimal, the difficult problem of assessing its merit arises.

The primary concern of this report is to consider the merits of the step-up procedure, to seek improvement in it and to investigate rules to govern the stopping of the selection procedure.

While this and related problems are of considerable interest and pertinence in the overall trajectory problem, they should not be considered overriding. Other approaches, where the goodness of approximation is more directly related to the cost criterion or to the equations of motion and where the mission fulfillment is more directly imposed, show at least equal promise and are being considered for subsequent study.

### Objectives

1. To conduct empirical investigation of the efficacy of using the step-up procedure in the selection of a fixed number of estimation variables out of a larger number in obtaining functional approximations by the method of LS.

2. To seek modifications of the procedure for the purpose of enhancing its efficiency.

3. To develop reasonable rules which will control the process of stopping the estimation variables selection procedure and to study empirically the sensitivity of the efficiency of the estimation to variations in these rules.

4. To explore empirically the general applicability of low-degree polynomial approximation (in the sense of least squares) to representative functions of several variables.

5. To develop an efficient, flexible and unified computer program which, in carrying out a least squares approximation, at least has the option of utilizing such selection procedures and stopping rules as have been developed.

#### Plan of Research

To accomplish the aims of this part of the study research was organized in four phases:

A. A review of the geometry, linear algebra and statistics involved in the method of least squares and the step-up procedure. This phase extended to include discussions of modifications to the step-up procedure and various criteria for stopping the selection process. Also included were algorithms for computer programs.

B. Development of the structure of the empirical investigations. In this phase decisions were reached on types of functions to be estimated, data patterns, size of data base, specific form of the estimation variables (as functions of independent variables), how data would be obtained and reduced to the regression format with particular regard to the important case of polynomial approximation.

C. Development of computer programs. In this phase algorithms developed in preceding phases were converted to programs, with attention to computational efficiency and cost.

D. A battery of examples with interpretations and, if possible, conclusions. In this phase a few preliminary examples were designed to test the efficiency of using the step-up procedure. Later, more sophisticated examples were used to develop the other objectives cited above.

#### Summary

- A. Mathematical review (see the supporting study titled: "Selection of Significant Estimation Variables in a Least Squares Problem: Mathematical Review.")

The well-known method of least squares (LS) is invoked to estimate a presumed functional relationship between a dependent variable  $Y$  and a set of independent variables  $X_1, \dots, X_{\pi}$  on the basis of a set of observed points. According to the method a class of functions of the form,

$$a_0 + a_1 Z_1(X_1, \dots, X_{\pi}) + \dots + a_p Z_p(X_1, \dots, X_{\pi}),$$

is considered for all real sets of coefficients. The  $Z$ 's are specified estimation variables depending on the independent  $X$ 's. For any function of the above class, corresponding to an observed vector of  $X$ 's, one could compute values  $z_{\mu 1}, \dots, z_{\mu p}$  of the estimation variables and a value  $\hat{y}_{\mu} = a_0 + a_1 z_{\mu 1} + \dots + a_p z_{\mu p}$ , which could be compared with the corresponding observed value  $y_{\mu}$  of the dependent variable  $Y$ . From this specified class of functions the method of LS selects one for which is minimized the sum of squares of the deviations of the so-called predicted values  $\hat{y}_{\mu}$  from the observed values  $y_{\mu}$ . Such a function is called a best estimate or best-fitting approximation (in the class) in the sense of LS.

The choice of the functions to be used as the estimation variables,  $Z_1, \dots, Z_p$ , is open, giving the method great flexibility, but also making it vulnerably dependent on the choice. In the next section of this summary some discussion is devoted to the choice of  $Z$ 's and the reduction of data to the form of observation vectors  $(y_{\mu}, z_{\mu 1}, \dots, z_{\mu p})$  on the variables  $(Y, Z_1, \dots, Z_p)$ . This form is now assumed.

The least squares approach admits of an accessible geometrical interpretation. Supposing there are  $N$  observation vectors, for each estimation variable  $Z_i$  consider the  $N$  observed values (adjusted to the mean). These values constitute the  $i$ -th estimation vector  $z_i$ . Similarly, consider the mean adjusted dependent-variable vector  $y$ . The LS problem translates to finding that vector in the space spanned by the estimation vectors which lies closest to the  $y$  vector. Or it may be interpreted as finding the projection of the  $y$  vector onto the estimation space.

The cosine of the angle between the  $y$  vector and its projection in the estimation space is called the multiple correlation coefficient,  $R$ . It is a measure of the efficiency of the estimate, attaining a maximum of unity when the  $y$  vector coincides with the projection estimate.

The difference between the  $y$  vector and its projection onto the estimation space is called the error vector. A pythagorean property holds, expressing the square of the length of the  $y$  vector as the sum of squares of the lengths of the estimate and the error. The estimate itself can be resolved into orthogonal components, and the same is true of the error vector.

If only  $k$  out of the  $p$  available estimation vectors are to be used to estimate  $y$  (corresponding to selecting  $k$  out of the  $p$  possible estimation variables), a difficult problem of deciding which  $k$  to elect arises, since trying all combinations is ordinarily computationally infeasible.

The step-up procedure is a practical, though not always perfectly optimal, way to select  $k$  estimation vectors. It evolves naturally from the geometric model described above. In this procedure the first estimation vector is chosen by finding the one on which the  $y$  vector has the longest projection (by the pythagorean property this leaves the shortest error vector). In the next step for each of the remaining vectors it is easy to determine the length of a component orthogonal to the first vector chosen, whose square added to the square of the projection of  $y$  on the estimation space of these two vectors. Selected is the vector having the longest such component. The procedure is then repeated.

Since the  $y$  vector may lie in the plane of two vectors but possibly closer to a third vector (not in the plane), the step-up procedure is not always optimal, for it would activate the third vector first, then one of the others, but the combination would not be as efficient as the first and second.

A modification of the procedure has been incorporated to allow for the elimination of a vector from the active estimation set. It works in



the following way. The error vector for the  $k$  selected variables is compared with the error vector when one vector is deleted from the active estimation set. The difference measures the net reduction of error due to the one vector deleted. Computationally it is easy to compare the lengths of these reductions. One may wish to eliminate a variable which contributes little net reduction. A measure of the net reduction due to each estimation vector is provided by the cosine of the dihedral angle formed by the plane containing the  $y$  vector and its projection in the reduced estimation space, on the one hand, and the plane containing the two projections, on the other hand. This is called the partial or net correlation coefficient between the dependent variable  $y$  and the estimation variable in question.

It appears evident that the simple rule of selecting  $k$  of  $p$  estimation vectors will not always be a good stopping rule. From the geometrical description several other natural criteria emerge as possible stopping rules whose use may be varied according to considerations of the particular problem at hand. For example, if the multiple correlation coefficient is "very high" the addition of other variables may seem unnecessary. Again, even if  $R$  is not high, the modified step-up procedure may be making no appreciable improvement in the estimate so that further addition of variables to the active estimation set may be deemed useless. Also, depending on the criteria for continuing to bring in new variables and to eliminate old ones, some stopping rule should be available to guard against cycling.

The most difficult choices for these decision rules are those concerning whether to eliminate an active estimation vector and whether adding one or several more will make any significant reduction in the error vector. One might adopt the rule of introducing two vectors and eliminating one, until a stopping rule stops the process. One might eliminate the vector to which corresponds the lowest net correlation coefficient, provided that the coefficient reaches a certain "low" value. One might stop adding vectors if the last  $r$  added make an average addition to  $R$  of less than some fixed amount. However, caution

should be exercised in the fixing of criteria, since certain combinations of these rules increase the chances of cycling.

Finally, we have considered elimination-stopping rule combinations based of F statistics. Briefly, an F statistic is a ratio of the average of certain of the estimation components to the average of the error components. In a statistical context, if the estimation components have on the average the same length as the error components, they are considered insignificant and are attributable to random error. In short these vectors are not considered of estimative significance. From such a point of view there is some intuitive appeal in the decision rule: Do not add if  $F \leq 1$ ; do not drop if  $F \geq 1$ . However, the rationale for using the F statistic rules is tenuous and, such as it is, depends on hypotheses of a statistical model which are not always appropriate. A fuller discussion of the statistical model is given in the supporting study.

While the mathematical and statistical analysis suggested the foregoing procedures and rules, it has also indicated considerable need for the empirical tests subsequently made.

The mathematical analysis included a translation of the geometrical steps described above into algorithms capable of being converted to computer programs. These well-known algorithms also are developed in detail in the supporting study with every effort made to retain geometrical interpretations in the development.

#### B. Structure of the Empirical Investigations

The data were organized in two main phases. The purpose of empirical runs in the first phase was primarily to gain insight on the efficiency of the step-up method for activating a subset of estimation variables out of a large set of such variables. The principal aim of the runs in the second phase was to explore the relative merits of various rules for stopping the step-up procedure of adding variables to the active estimation set and rules for eliminating such variables. Auxiliary purposes of empirical runs were to test and correct pertinent computer programs and to obtain from diversified experience an idea of the general validity of the LS approach as an approximation technique.

As pointed out in the previous section, the generality of the method of LS leaves considerable latitude in the selection of test cases. In organizing test runs representing a variety of problem types some of the factors on which decisions had to be reached included:

1. The type of function to be approximated, including its form, the number of variables and the selection of a representative member.
2. The class of approximating functions, i.e., a selection of the estimation variables  $Z_i = Z_i(X_1, \dots, X_n)$ ,  $i = 1, 2, \dots, p$ , where  $(X_1, \dots, X_n)$  presumably is in the domain of the function to be approximated.
3. The number, extent and distribution of data points.

Admittedly decisions reached during the test construction concerning these factors were somewhat arbitrary. They were made, however, with awareness of their significance.

Briefly, it was decided to construct data for a few selected functions of three variables, using a rectangular grid of data and balanced polynomials as approximating functions. In addition, a few runs were made using active data, which were developed in certain statistical regression analyses. Except for the actual data runs the data grids consisted of 500 to 1000 points generated from evenly spaced values of the three variables on the margins. Thus the undoubtedly important effects (on goodness of fit) of varying the distribution of data points or varying the types of estimating functions were not studied here. Indeed these factors were held more or less constant in order not to obscure the comparisons of variable-selection procedures.

These decisions led to fairly general and easy algorithms for generating data for a given test run and reducing them to the format of LS input. Thus, for a given function  $F(X_1, X_2, X_3) = Y$ , a given class of balanced polynomials of the form

$$\tilde{Y} = \sum a_{\ell_1 \ell_2 \ell_3} X_1^{\ell_1} X_2^{\ell_2} X_3^{\ell_3},$$

and a given rectangular grid of points,

$$(x_{1t_1}, x_{2t_2}, x_{3t_3}),$$

observation vectors  $(y_\mu, z_{\mu 1}, z_{\mu 2}, \dots, z_{\mu p})$  were generated by the computer. Here  $y_\mu$  is the value of  $Y$  at some  $(x_{1t_1}, x_{2t_2}, x_{3t_3})$ , and the estimation

variables  $Z_i$  are the several terms of the balanced polynomial of the form

$$Z_i = X_1^{\ell_1} X_2^{\ell_2} X_3^{\ell_3},$$

while  $z_{\mu i}$  is the value of  $Z_i$  when  $(X_1, X_2, X_3) = (x_{1t_1}, x_{2t_2}, x_{3t_3})$ . The observation vectors were then in a form to obtain LS estimates of the coefficients in the best-fitting balanced polynomial, or more specifically to manipulate in a way aimed at activating the most significant estimative terms of the balanced polynomial as described in the foregoing section.

Runs in the first phase were limited to estimating a polynomial (of higher order than the approximating ones) and estimating a rational function, while the approximating balanced polynomial class was restricted to be of second degree in  $X_1$  and  $X_2$  and first degree  $X_3$ , which restricted the number  $p$  of estimation variables (terms of the polynomial) to 17 or less. The test procedure for these runs was, for each  $k = 1, 2, \dots, p-1$ , to determine the efficiency (multiple correlation) of each of the  $\binom{p}{k}$  subsets of  $k$  vectors and compare the optimal set with the set produced by the step-up procedure. Computer time was a limiting factor in these tests.

Runs in the second phase included estimating an exponential function and a few algebraic functions other than rational functions, and they included two runs using actual statistical data. Some effort was made to include poorly fitted functions as well as accurately fitted ones. Also, the form of the approximating balanced polynomial was stepped up to develop 47 estimation variables. Usually, for each example, several runs were

initiated in which were varied the policies of stopping the selection procedure or of eliminating a variable.

Considered, but not developed in this study, was an experimental design in which runs would be made for the various different combinations of prescribed levels of the main factors thought to influence efficient variable selection.

C. Development of Computer Programs (see the supporting study titled, "Selection of Significant Estimation Variables in a Least Squares Problem: Computer Programs.")

Corresponding to the two phases of the study mentioned in the last section, two computer programs were developed. The purpose of the first program was to compare in a few examples the subset of  $k$  estimation vectors selected by the step-up procedure with the optimal subset of  $k$ . This first phase of programming was begun before the Burroughs 5000 was operational on contractor facilities and was programmed in the ALGOL 58 compiler language for the Burroughs 220 computer. Because of core memory limitations the program restricts the total number of estimation vectors to twenty-five. It would be a simple matter to translate the program to one for the more advanced computer. This has not yet been done, primarily because the number of comparisons to be made even with the restriction to 25 variables makes for an almost prohibitive amount of computation time.

The program depends on using (1) rectangular grid data and (2) a balanced polynomial as the general form of the approximating function. One part of the program, using as input the specified values of each of the variables and the degree of the balanced polynomial in each variable, generates internally the grid of data points and computes for each such point the value of each term of the balanced polynomial. Thus the estimation vectors are generated.

Also the program allows for a procedure to be inserted to incorporate the computation of the values of the function which is to be approximated, at each of the grid points of data. Thus the dependent variable vector  $y$  is generated.

As an intermediate calculation the program mean adjusts the above vectors and produces the intercorrelation matrix for all the vectors, including the dependent variable vector. There will be  $L_1 L_2 \dots L_{\pi} = p + 1$  such vectors. These are restricted in number to 25.

In the next part of the program, for each  $k = 2, 3, \dots, p-1$ , each one of the  $\binom{p}{k}$  subsets of  $k$  estimation vectors is manipulated to compare the estimation efficiency (multiple correlation) of those subsets. For each  $k$  the subset of  $k$  vectors which gives maximum efficiency is printed as is also its corresponding multiple correlation coefficient.

In the final part of the program the estimation vectors are selected in the order prescribed by the step-up procedure. At each stage an index of the estimation vector introduced at that stage is printed out, as well as the multiple correlation coefficient obtained with the set of vectors selected up to that stage.

In this program checks were instituted to restrain the incorporation of vectors which were practically dependent on vectors already included in the active estimation set. Also, considerable effort was made to abbreviate the matrix-inversion type calculations in order to produce only the multiple correlation, since the number of such calculations,  $2^p - p - 2$ , rapidly gets large.

The purpose of the second program, to a considerable extent based on the assumption that the step-up procedure was reasonably efficient, was to make available a fairly flexible program for estimations based on the method of LS in which would be included at least options for activating subsets of the estimation variables according to the step-up procedure and other modified procedures, and also included would be options which could be exercised to stop the selection. The program was done for the Burroughs 5000 in the ALGOL 60 compiler language.

As it now stands the program has several options for obtaining the basic matrix of the dot products of the adjusted vectors (which matrix reduces to the intercorrelations matrix when the rows and columns are appropriately standardized).

- (1) One of these options is the same as in the previous program, except that the admissible order of the matrix has now been increased to more than 100. This option allows for the rapid generation of data for experimental studies.
- (2) Either the matrix of dot products or the intercorrelation matrix may be read in directly. This allows further study, especially of subset selection procedures, of previously studied regression problems, least squares fittings, and so forth.
- (3) Observation vectors may be directly read in. This will be the way data will arise in most realistic problems, although values of the estimation variables may require preliminary transformation (e.g., if the estimation variables are terms in a balanced polynomial).

In this program, once the basic matrix has been obtained, it is retained in memory and can be used over and over, to facilitate comparisons when various procedures for selection, elimination, and stopping are employed.

In case the intercorrelations matrix was not introduced directly the program gives an option for computing and printing it and using it in the remainder of the program.

In the main part of the program estimation vectors are introduced in the priority order dictated by the step-up procedure. In addition, however, the procedure carries options which allow for various rules to be set to make possible the elimination of an estimation vector and the stopping of the selection process.

At present there are two criteria either one of which may be used to eliminate an estimation variable. One option automatically eliminates an estimation variable after two have been included. Of course the one deleted is the one of lowest net correlation with the dependent variable (see Section A preceding). In the other option the pertinent F statistic for the variable with smallest net correlation is computed (see Section A)

and is tested against a preassigned threshold value. If it is below this value, the variable is deleted. It is possible to prevent any such eliminations by setting the threshold equal to zero.

Currently there are four criteria which can be used to stop the process of adding estimation variables. The program effectively permits bypassing any or all of these criteria. They are:

- (1) Stop if the F ratio for the next single variable to be introduced does not exceed that threshold value corresponding to a preassigned significance level. The procedure stops after that estimation variable has been added. This can be bypassed by setting the threshold at zero.
- (2) Stop if the current value of the multiple correlation coefficient is sufficiently large. This can be bypassed by setting the multiple correlation threshold at unity.
- (3) Stop if the number of variables chosen reaches a preassigned number. This can be bypassed by setting that number equal to the total number available.
- (4) Stop when the number of computational iterations for adding or eliminating a vector has exceeded a preassigned number.

It is noteworthy that the computational procedures for eliminating and for adding a vector are the same, once the vector has been earmarked.

It should also be mentioned that the same precautions as in the earlier program were taken to prevent the introduction of almost linearly dependent vectors.

In this program of course the output includes the LS regression coefficients of the selected estimation vectors, as well as indices of the vectors selected, and the multiple correlation coefficients.

- D. Test Runs on the Computer (see the supporting study titled, "Selection of Significant Estimation Variables in a Least Squares Problem: Empirical Computer Studies.")

As indicated in previous sections, these tests were broken roughly into two phases. In a very limited way the preliminary set of tests was



conducted to gain a measure of confidence in the step-up procedures as a means for selecting an efficient subset of estimation variables in a least squares model. In the tests made a balanced polynomial of relatively low order was selected, the terms of which provided the full set of estimation variables. Estimation vectors, as well as a dependent-variable vector, were generated from rectangular design data. Dependent-variable data were computed as values of the function which was to be approximated. As described previously, subsets of estimation vectors selected by the step-up procedure were compared with the optimal set. Primary difficulty in test runs arose from fact that the determination of the actual optimal set of  $k$  vectors required comparisons of  $\binom{p}{k}$  sets of vectors, where  $p$  was total number of estimation variables available. Computational feasibility dictates that  $p$  be severely restricted.

Nevertheless, several preliminary runs were made where  $p$  was kept to about 11, and in all cases less than 18. Several functions were approximated. These in general represented the class of rational functions. For one of the functions, which had a pole in the region of data points, only a poor approximation was obtained. Otherwise, even with low-degree polynomials, the multiple correlation coefficient was rather high.

In most of these tests the step-up procedure selected, at each stage, the optimal set of variables. There was one example, however, where the procedure did not select the optimal set of two vectors, although the correct selection of a larger number of variables was achieved. It is also noted that, when  $R$  became stable or nearly so, additional variables introduced by the step-up procedure were not always optimal. It is possible that this could have been the result of round-off error.

In general these experimental results indicated the step-up procedure is probably quite efficient, at least when a fair scatter of points is available. It was noted that, even when the method failed, the value of  $R$  was near optimal. The actual occurrence of failures, even at early stages, suggested that some means for eliminating variables would be desirable. Such techniques were introduced and used in the second phase of testing.

For the second set of test runs the Burroughs 5000 program was used. As mentioned earlier, this program allows for a larger number of estimation vectors to be handled, incorporates options of data input, variable elimination and program stops, but does not make the comparisons to determine a purely optimal subset of estimation variables. In most of the examples studied in this phase several runs were made for each example to throw light on the effects of changing the pattern of variable elimination and stopping rules. Attention was focused on varying the elimination criterion, the effects of varying other rules being discernible from the print-out, with the principal basis for elimination being an F statistic (see Section A of Summary). To observe the effect of certain stopping rules (which can be set in the program options) print-out includes for each "sweep" (where a variable is eliminated or added to the estimation set) the number of sweeps up to that stage, the number of estimation variables being used, an index of the last one eliminated or added, the  $F_I$  value of the F statistic for a variable brought in or the  $F_0$  value of the F statistic corresponding to a variable being eliminated (if it was below the criterion level), and the square  $R^2$  of the multiple correlation coefficient, as well as the reduced  $R^2$  which diminishes if and only if the last variable introduced gave an  $F_I$  value less than unity.

The examples included: Approximating three non-polynomial functions, with the available variables being the 48 terms of a balanced polynomial cubic in  $X_1$  and  $X_2$  and quadratic in  $X_3$  and the 500 data points generated from  $X_1 = 0.25(0.25)2.50$ ,  $X_2 = 0.25(0.25)2.50$  and  $X_3 = 0.25(0.25)1.25$ ; approximating a dependent variable from actual data with available variables constituting a balanced polynomial in four variables, where the data are (as would usually be the case in practice) not in rectangular design; and approximating a dependent variable from actual data where the intercorrelation matrix of available estimation vectors was given, the presumption being that these could be non-polynomial terms.

In the first group of examples the functions chosen to be approximated were

$$F_1(X_1, X_2, X_3) = \exp(-X_1^2 X_2 X_3)$$

$$F_2 = (X_1^4 + X_2^3 + X_3^2) | X_1 + X_2 - \frac{\pi}{2} X_3 |^{-\frac{1}{2}}$$

$$F_3 = \sqrt{X_1^2 + X_2^2 + X_3^2}.$$

As in all examples the data were mean-adjusted. The functions  $F_1$  and  $F_3$ , especially  $F_3$ , were very closely approximated (in the range of data) by the full set of 47 estimation vectors in the sense that  $R^2$  was near unity, while  $R^2$  for the case of  $F_2$  was near 0.9. For each example runs were made with  $F_0$  set over a range of values from high to low. In the case where  $F_0$  was set very low the tendency was to eliminate few or no variables and thus to be very close to the simple step-up procedure.

The test runs for these examples show that different subsets of estimation variables will be selected when the elimination (and stopping) rules are varied. They provide concrete examples wherein the step-up procedure is bettered by a procedure modified to include an elimination criterion; where the opposite happens; where an  $F_1$  stopping criterion of 1.00 (on the last variable brought in) could stop the procedure which if continued would later introduce variables significant at this same level. These test runs suggest, but not markedly or universally, that the elimination criterion is effective in obtaining a higher  $R^2$  for the same number of estimation variables; that a high criterion value is more effective for variables selected early but not for those selected later; that the  $F_1$  test may stop the procedure too soon unless modified; that different problems seem to need somewhat different rules; that while the set of variables selected may vary considerably  $R^2$  has a tendency to be fairly stable for different procedures.

The examples with actual data provided experience with data more of the type expected in a realistic problem. In addition the first provided a good example in which an F stopping rule based on a single variable (last introduced) would have stopped the procedure too soon. The last example illustrates another point, viz. that out of 14 variable the last nine variables tested together are not significant at 50% level while the 6th one tested alone is significant at this level.

It should be noted that in all the examples, in terms of the multiple correlation coefficient, a few estimation variables usually accounted for most of the value of  $R^2$ .

It is recommended that further insight be obtained by examining the summary data for the various test runs, given in the supporting study referred to above.

#### Conclusions and Recommendations

The step-up procedure, which first activates the one estimation variable best in the sense of least squares, activates next the one which contributes the most to a further reduction in the sum of squares and so forth, is supported as an efficient and computationally feasible procedure for selection priority-rated estimation variables in a least squares approximation problem.

The nonoptimality of the procedure is manifest in practice. However, the evidence is strong that even in such case the results are near-optimal, as measured by the multiple correlation coefficient,  $R$ . The empirical evidence indicates more reliability of the step-up procedure in the activation of the earlier and presumably more significant variables than in later variables. When a large number of estimation variables is involved, the optimal value of  $R$  appears to be nearly reached by several subsets of estimation vectors. Thus, although frequently in these cases the set selected by the step-up procedure is not optimal, it is very nearly so.

If it is important to restrict the number of estimation variables, there appears to be a need for a means of eliminating variables previously activated. The procedure of eliminating an active variable whose net contribution to the reduction in the sum of least squares is (and small) is practicable and frequently effective. Examples show, however, that the elimination modification does not always improve on the simple step-up procedure. Moreover, it carries the same cost as activating an

estimation variable. No fixed elimination criterion is best for any wide variety of problems. The experiments indicated an overall tendency for a large elimination criterion to be more effective when the active estimation subset is small and a small criterion to be more effective when the number of active estimation variables has become sizable.

The use of rules to stop the activation of additional estimation variables must often depend on such factors as available computer time and rate of computer time utilization. A comprehensive set of rules, which may be used in various combinations, includes stopping when R is sufficiently large, when the activation of additional variables does not contribute significantly to the estimation, when the number of variables reaches a preassigned number or when the computational procedure begins to cycle. Examples show that the second of these can occasionally stop the process too soon, so that the contribution of the last several active variables, rather than just the last one, should probably be tested. The speed with which variables were eliminated or introduced in the examples indicates that large blocks of variables could be introduced before making any decision on which variables to keep active.

The study shows that at the current state of computer science it is still infeasible to examine all combinations of subsets of estimation variables to determine the optimal subset, unless the total number is quite small, and thus that the need remains for such a procedure as the step-up procedure. The study has also given evidence of the feasibility of the rapid selection of efficient estimation variables even from a set of several hundred, using a fairly sophisticated system of optional variable-elimination and stopping rules.

Finally, with reservation, it should be noted that in all the examples there was a marked relative efficiency of a small set of active estimation variables to the entire set of estimation variables available.

In view of the foregoing results the step-up procedure is recommended as an effective means for selecting priority-rated estimation variables

in a least squares analysis. The use of the modified procedure and the various stopping rules is also recommended with the admonition that the various settings ought insofar as possible to be adjusted to suit the experience of workers familiar with the problem area under study.

Specifically, with regard to the context of estimating optimal trajectories, i.e. with regard to the problem giving rise to this study, it is recommended that further general analysis of the method described herein, either theoretical or empirical, not be undertaken, but that the method and experience gained be applied in a series of experiments with actual trajectory data as soon as possible, where the experience of researchers in the field and the knowledge of physics pertinent to the problem will be utilized to help delimit the class of approximating functions.

Finally, using methods of design of experiments and a limited class of functions presumably pertinent to trajectory problems and including some live data, it may be feasible to study the effects (on approximation efficiency) of varying certain factors such as data distribution, type of approximation, elimination criterion, and so on.

# SELECTION OF SIGNIFICANT ESTIMATION VARIABLES IN A LEAST SQUARES PROBLEM: MATHEMATICAL REVIEW

1. Introduction. The principle of least squares (LS) can be formulated in the following terms. Presumed to exist is some sort of functional dependence of one variable,  $Y$ , called the dependent variable, on a vector,  $(X_1, \dots, X_\pi)$ , of  $\pi$  other variables, called independent variables. Available is a number (say  $N$ ) of observations, i.e. values of  $Y$  corresponding to values of the vector  $(X_1, \dots, X_\pi)$ . Next is chosen a class of admissible functions of the form,  $a_1 Z_1(X_1, \dots, X_\pi) + \dots + a_p Z_p(X_1, \dots, X_\pi)$ , where the  $Z_1, \dots, Z_p$  are fixed functions of the  $X$ 's and the parameters of the class are  $a_1, a_2, \dots, a_p$ . The functions  $Z_i$  presumably are chosen to enhance the likelihood that the unknown functional relationship (between  $Y$  and the  $X$ 's) will be nearly of the prescribed form. Each function of the class is linear in the variables,  $Z_1, \dots, Z_p$ , which we shall call estimation variables; each function is also linear in the parameters. In any case the basic idea in the least squares approach is to approximate the unknown functional relationship with one of the admissible functions. For any one of the functions in the class, corresponding to each observation,  $(x_{\mu 1}, \dots, x_{\mu \pi})$ , is the value of the function,  $\tilde{y}_\mu = a_1 z_{\mu 1} + \dots + a_p z_{\mu p}$ , where  $z_{\mu i} = Z_i(x_{\mu 1}, \dots, x_{\mu \pi})$ , which is comparable with the value of  $Y$  (say  $y_\mu$ ) corresponding to this same observation,  $(x_{\mu 1}, \dots, x_{\mu \pi})$ . The sum of squares,

$$\sum_{\mu=1}^N (\tilde{y}_\mu - y_\mu)^2,$$

is taken as a measure of the estimative value of the function  $\tilde{Y} = a_1 Z_1 + \dots + a_p Z_p$ . According to the principle of least squares, out of the class of admissible functions

$$y = \{\tilde{Y} | \tilde{Y} = a_1 Z_1 + \dots + a_p Z_p\}$$

is chosen as an estimate of  $Y$  one function which minimizes the sum of squares of deviations. Such an estimate (we shall see that one does exist) is written as  $\hat{Y} = \sum_1^p b_i Z_i$ ; we shall call such a function a best estimate or best-fitting approximation (in the class) in the sense of least squares. The sum of squares of deviations,  $\sum_1^N (\hat{y}_\mu - y_\mu)^2$ , is called the sum of least squares or the residual sum of squares due to error. The procedure of obtaining a best estimate in the above sum is frequently called a regression analysis, or more properly a linear regression analysis. The  $b_i$  are often called regression coefficients.

The method of LS was known and used by Gauss over 150 years ago. He discovered that under certain conditions the method of least squares in a sense yields an optimal estimate. This is the famous Gauss-Markov theorem. Briefly, the principal hypothesis for this theorem is that except for random deviations the observed values of  $Y$  are values corresponding to one of the functions in the class  $\mathcal{Y}$ . The random deviations are assumed to be statistically uncorrelated, with a common variance and mean zero. Under the additional hypothesis of normality of the distribution of these deviations an elegant statistical theory of estimation and hypothesis testing can be constructed. The statistical model is discussed briefly in Section 5 below.

The method of LS is used widely in numerical analysis even when the support of the Gauss-Markov theorem cannot honestly be invoked. In many cases other methods perhaps are equally or more justifiable; but often the method of LS has an intuitive appeal in that it seeks an estimate which minimizes one obvious measure of error.

It is also possible to consider classes of admissible functions, from which an estimate will be chosen on the basis of the LS principle, which classes are nonlinear in the parameters. In many instances such problems are resolved satisfactorily by iterative techniques. The procedure of obtaining estimates of the parameters in such a case is called a nonlinear regression analysis.



Excellent accounts of the statistical linear regression model are given in GRAYBILL, SCHEFFE, and ZELEN. The method of LS is given space in most numerical analysis books, and sometimes the nonlinear case is discussed. E.g., see SCARBOROUGH. Nonlinear regression analysis is treated from a statistical point of view in WILLIAMS.

In applications of LS it is often the case that the number of estimation variables, for which values are computable from observations on independent variables, is very large. Certain recurring and nagging questions arise, varying somewhat with the circumstances. If only  $k$  of  $p$  variables can be used, which  $k$  should be chosen? Does the use of additional variables contribute significantly to increased efficiency of estimate? The second of these questions is not mathematically meaningful until the word "significantly" is defined. However, in the context of a given problem, the question is one that frequently must be raised, given meaning and acted on.

There is an obvious answer to the first question raised above, viz., to determine by computation which of the  $\binom{p}{k}$  sets of estimation variables yields the minimum sum of least squares from the data. Unfortunately this straight-forward procedure is computationally infeasible. A more tractable and completely reliable method of finding the optimal set of  $k$  estimation variables remains an open problem. However, at least as early as 1931, WHERRY proposed a procedure for selecting a reasonably efficient subset of estimation variables. This procedure we call -- because it has become our habit -- simply the step-up procedure. It consists in selecting first the one estimation variable best in the sense of LS, next the one which contributes the most to a further reduction in the sum of LS, and so forth. In this way variables are added until some rule stops the process. The procedure is computationally very feasible and fast. However, it is easy to show it is not always optimal. The step-up procedure has recently been described without much critical analysis in papers by ANDERSON and FRUCHTER, and SCHULTZ and GOGGANS.

The aims of the present paper are: To illuminate the method of LS in linear regression analysis with geometrical arguments, giving clear interpretation of certain measures of estimation efficiency; thus to lead into a natural development of the step-up procedure where its weakness as well as its intuitive appeal are exposed; to examine the geometrical structure for a procedure for elimination of a variable previously selected, and thus mitigate the flaws in the step-up procedure; to explore the statistical model for reasonable decision rules on when to eliminate and when to keep adding variables; and finally to provide a translation of the various geometrically conceived procedures to computable algorithms.

2. Geometric formulation of the principle of least squares. The notion of obtaining an estimate,  $\hat{Y} = \sum_1^p b_i Z_i$ , out of the admissible class which minimizes the sum of squares of deviations, is one admitting of accessible and correct geometrical descriptions. Such a formulation is helpful in understanding the step-up procedures for selecting significant estimation variables (to be described in the next section) and seems to hold the only hope of devising techniques even more defensible than the step-up procedure. We proceed now toward such a formulation.

Assumed available are the  $N$  observation vectors,  $(y_\mu, z_{\mu 1}, \dots, z_{\mu p})$ ,  $\mu = 1, 2, \dots, N$ , where  $z_{\mu i} = Z_i(x_{\mu 1}, \dots, x_{\mu n})$ , as indicated in the preceding section. Associated with each of the  $p$  estimation variables  $Z_i$ ,  $i = 1, 2, \dots, p$ , is the vector, lying in the euclidean  $N$ -space  $E^N$ , consisting of  $N$  values  $z_{\mu i}$ ,  $\mu = 1, 2, \dots, N$ , observed on that variable. We shall call these vectors estimation vectors; we write them,  $z_i$  ( $i = 1, 2, \dots, p$ ); and for matrix manipulations they will be thought of as column vectors. Hence, using the letter  $T$  to indicate matrix transpose,  $z_i^T = (z_{1i}, z_{2i}, \dots, z_{Ni})$ . In this section the  $N \times p$  matrix of these estimation vectors will be denoted as  $z$ . Similarly, the symbol  $y$  represents the vector of the observed values of the dependent variable  $Y$ . It will be assumed, without any real loss of generality, that  $N > p$  and that the estimation vectors are linearly independent. Thus the estimation vectors constitute a basis of a  $p$ -dimensional vector space  $V_p$ , lying in  $E^N$ .

Consider now the sum of squares criterion. Writing the parameter vector as  $a$ , this criterion is

$$g(a) = \sum_{\mu=1}^N (y_{\mu} - \tilde{y}_{\mu})^2 = d^T d,$$

where  $d = y - \tilde{y}$  is the vector of deviations. Note that  $\tilde{y} = \sum_1^p a_i z_i$  lies in the vector space  $V_p$  generated by the estimation vectors and that  $d^T d$  is the square of the (euclidean) distance between  $y$  and  $\tilde{y}$ . Since the aim was to determine  $b$  such that  $g(b) = \min \{g(a) | a\}$ , the least squares problem may be interpreted as finding a vector in the space spanned by the estimation vectors which lies nearest the dependent-variable vector  $y$ .

Geometrical intuition now supplies the correct solution to the least squares problem; viz., the vector in  $V_p$  lying nearest  $y$  is the projection of  $y$  onto  $V_p$ . Other important points are indicated by the geometry. Writing  $\hat{y}$  as the projection of  $y$  onto  $V_p$ ,  $e = y - \hat{y}$ , and  $e^2 = e^T e$ , etc., pythagorean relations are indicated. E.g.,  $y^2 = \hat{y}^2 + e^2$ ; i.e., the square of the length of the dependent-variable vector equals the sum of the squares of the lengths of the best estimate vector and the least squares residual error vector. This is often stated as, "The total sum of squares equals the sum of squares due to regression (estimation) plus the sum of squares due to error." Also, if  $\tilde{y} = \sum_1^p a_i z_i$  is another vector lying in  $V_p$ , if  $d = y - \tilde{y}$ , then  $d^2 = e^2 + (\hat{y} - \tilde{y})^2$ . Also, the  $e$  vector will be orthogonal to  $V_p$ . Finally, the angle between  $y$  and its projection should be less than the angle between  $y$  and any other vector in  $V_p$ . Thus  $\cos \theta(y, \hat{y}) > \cos \theta(y, \tilde{y})$ , where  $\theta(u, v)$  means the angle between vectors  $u$  and  $v$ .

In statistical terminology the cosine of the angle between two such vectors is called a correlation coefficient. Recall that

$$\cos \theta(u, v) = \frac{(u \cdot v)}{\sqrt{u^2 v^2}} = \frac{\sum_1^p u_i v_i}{\sqrt{\sum_1^p u_i^2 \sum_1^p v_i^2}}$$



The foregoing geometrical discussion can be substantiated with a detailed algebraic development. Such substantiation is a consequence of the argument to follow, but the primary purpose of the argument is to make the geometrical entities explicit, to make essential quantities computable and to set the stage for the next section.

The estimation space  $V_p$  is spanned by sets of orthogonal vectors of unit length. Let  $z_1^*, z_2^*, \dots, z_p^*$  be one such set. Since every vector in  $V_p$  is a unique linear combination of the estimation vectors,

$$z_1^* = q_{11}z_1 + \dots + q_{1p}z_p$$

$$\vdots \quad \vdots \quad \vdots$$

$$z_p^* = q_{p1}z_1 + \dots + q_{pp}z_p$$

i.e.,  $z^* = zQ$ , where  $Q$  is a non-singular  $p \times p$  matrix, and, of course,  $z = z^*Q^{-1}$ . Also, every vector in  $V_p$  has a unique representation either as a linear combination of  $z_1, \dots, z_p$  or of  $z_1^*, \dots, z_p^*$ . If  $\tilde{y}$  lies in  $V_p$ , then there exists a unique vector  $a$  such that  $\tilde{y} = \sum_1^p a_i z_i = za$ , and there exists a unique vector  $a^*$  such that  $\tilde{y} = z^*a^*$ . But  $z = z^*Q^{-1}$ , so that  $a^* = Q^{-1}a$ . Thus there is a one-one correspondence between coefficient vectors  $a$  for the  $z$  basis and vectors  $a^*$  for the  $z^*$  basis. In particular, if  $b^*$  is such that  $\hat{y} = z^*b^*$  is the one vector in  $V_p$  closest to  $y$ , then  $\hat{y} = zb$ , where  $b = Qb^*$ .

With these orthogonal vectors  $z^*$  in mind an orthogonal transformation is now imposed on the points in  $E^N$  in such a way that, in the transformed space  $E^{N'}$ , the  $z^*$  become the unit vectors  $u_1, u_2, \dots, u_p$ . Such a transformation is accomplished with an  $N \times N$  orthogonal matrix  $P$  whose first  $p$  rows are the vectors  $z^{*T}$ . It is easily seen that distances and angles are preserved under such a transformation, so that the least squares problem is invariant under the transformation. Note that the image  $V_p'$  of  $V_p$  is simply the linear combinations of the unit vectors,

$u_1, \dots, u_p$ . Let  $y' = Py$ , and let  $\tilde{y} = z^*a^*$  lie in  $V_p$ , so that  $\tilde{y}' = \sum_1^p a_i^* u_i$ .

Then the square of the error vector is

$$d^T d = d'^T d' = (y' - \sum_1^p a_i^* u_i)^T (y' - \sum_1^p a_i^* u_i) = \sum_1^p (y_i' - a_i^*)^2 + \sum_{p+1}^N y_{\mu}'^2.$$

Evidently the projection of  $y'$  onto  $V_p$  ought to be the vector whose first  $p$  components are those of  $y'$  and whose remaining components are zero.

Thus the  $a_i^*$  which produce the combination of  $u_i$  ( $i = 1, 2, \dots, p$ ) constituting the projection of  $y'$  on  $V_p$  are  $y_i'$ . In short,  $b_i^* = y_i'$ ,  $i = 1, 2, \dots, p$ .

That this is correct algebraically can be seen in the preceding equation,

where it is obvious that these are the values of  $a_i^*$  which minimize the square of the error vector. Write  $\hat{y}' = \sum_1^p b_i^* u_i = [y_1', \dots, y_p', 0, \dots, 0]^T$ .

Note that the residual error vector  $[0, \dots, 0, y_{p+1}', \dots, y_N']^T = e'$  so that

$e'$  and  $\hat{y}'$  are orthogonal. Note also that  $\sum_1^p (y_i' - a_i^*)^2 = (\hat{y}' - \tilde{y}')^2$  and hence, from the foregoing equation, that

$$d'^2 = (\hat{y}' - \tilde{y}')^2 + (y' - \hat{y}')^2 = (\hat{y}' - \tilde{y}')^2 + e'^2.$$

Having seen now that, relative to an orthogonal basis of  $V_p$ ,  $b^* = z^{*T}y$  (which follows from the fact that  $b_i^* = y_i'$  and  $y_i' = z_i^{*T}y$  for  $i = 1, 2, \dots, p$ ), it is now desirable to obtain  $\hat{y}$  and  $e$  in terms of the original estimation vectors and the dependent variable vector. But  $\hat{y} = z^*b^* = zb$ , where  $b = Qb^* = Qz^{*T}y = QQ^T z^T y$ . Now

$$(QQ^T)^{-1} = Q^{-1T} Q^{-1} = Q^{-1T} z^* z^{*T} Q^{-1} = (z^* Q^{-1})^T (z^* Q^{-1}) = z^T z.$$

Thus, writing  $h = z^T z$  and  $g = z^T y$ , in terms of original data,  $b = h^{-1}g$ .

Also

$$e^2 = \sum_1^p y_i'^2 = b^{*2} = y^T z^* z^{*T} y = y^T z Q Q^T z^T y = (z^T y)^T b = \sum b_i g_i.$$

Thus computationally the problem is one of solving the system of equations  $hb = g$ . In the succeeding discussion it will be important to remember the following principle which summarizes much of the preceding development and unifies the geometry and algebra of the least squares problem: Given a set of  $k$  linearly independent vectors  $z_1, \dots, z_k$  in an euclidean space and a  $(k+1)$ -st vector  $w$ , if  $h = z^T z$  where  $z = (z_1, \dots, z_k)$  and  $v = z^T w$ ; then the solution  $x$  of the equations  $hx = v$  is such that  $zx$  is the projection of  $w$  onto the space generated by the  $z_i$ , and the solution effectively resolves the  $w$  vector into its projection  $zx$  and a component,  $e = w - zx$ , orthogonal to the projection.

3. The Step-up Procedure. In this section emphasis is shifted to the selection of a subset of (say)  $k$  estimation vectors out of a total number of (say)  $p$ . An optimal set of  $k$ , by definition, will be that set of  $k$  corresponding to which the length of the error vector is least (or equivalently the multiple correlation coefficient  $R$  is most). The plausibility of the step-up procedure, as well as its deficiencies, will be seen from the geometrical development. Computational feasibility and procedures will be evident from the corresponding algebra.

For the moment we suppose that  $k-1$  vectors have been chosen and that our purpose is to add another one from the  $p-(k-1)$  remaining. We shall refer to estimation vectors selected as being in the active estimation space or as being active.

With regard to a least square problem involving  $y$  and the  $k-1$  active estimation vectors (which of course are a basis for a vector space  $V_{k-1}$  of dimensionality  $k-1$ ) everything in the preceding section is directly applicable. This succession of problems with  $1, 2, \dots, k, \dots, p$  vectors in the active estimation space is sometimes called the succession of the 1st, 2nd, ...,  $k$ th, ...,  $p$ th  fittings . We shall frequently use a superscript to indicate the fitting, or dimension of the active estimation space. This notation does not specify which of the vectors are in the active estimation space, but we shall tacitly assume they have been re-labeled so that the active estimation vectors are now  $z_1, z_2, \dots, z_{k-1}$ .

According to the preceding section  $\hat{y}^{(k-1)} = \sum_{i=1}^{k-1} b_i^{(k-1)} z_i = z^{(k-1)} b^{(k-1)}$ , where  $b^{(k-1)}$  is the solution to the system of equations,  $h^{(k-1)} b^{(k-1)} = g^{(k-1)}$ , with  $h^{(k-1)} = z^{(k-1)T} z^{(k-1)}$ ,  $g^{(k-1)} = z^{(k-1)T} y$ , and  $z^{(k-1)} = (z_1, \dots, z_{k-1})$ . Recall that  $\hat{y}^{(k-1)}$  is the projection of  $y$  onto  $V_{k-1}$  and that the residual error vector  $e^{(k-1)}$  has length whose square is  $(b^{(k-1)} \cdot g^{(k-1)})$ .

Suppose next that the  $k$ th vector to become active has been selected. Consider the system of equations  $h^{(k-1)} x^{(k-1)} = v^{(k-1)}$ , where  $v^{(k-1)} = z^{(k-1)T} z_k$ . Recall that  $\sum_{i=1}^{k-1} x_i^{(k-1)} z_i$  is the projection of  $z_k$  onto  $V_{k-1}$ , and  $z'_k = z_k - \sum_{i=1}^{k-1} x_i^{(k-1)} z_i$  is the component of  $z_k$  lying orthogonal to the space spanned by the  $z_1, \dots, z_{k-1}$ . The vectors  $z'_1 = z_1, z'_2, \dots, z'_k, \dots$ , thus defined are a particular determination of Gram-Schmidt orthogonal vectors. In matrix form the matrix of the first  $k$  of these Gram-Schmidt vectors is

$$z'^{(k)} = z^{(k)} Q'^{(k)}, \text{ where } Q'^{(k)} = \begin{bmatrix} 1 - x_1^{(1)} - x_1^{(2)} & \dots & -x_1^{(k-1)} \\ 0 & 1 & -x_2^{(2)} & \dots & \vdots \\ \vdots & 0 & 1 & \dots & \vdots \\ & \vdots & \ddots & \ddots & -x_{k-1}^{(k-1)} \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix}$$

from the equation above.

Normalized Gram-Schmidt vectors are obtained when the columns of  $Q'^{(k)}$  are divided by  $(z'_i \cdot z'_i)^{\frac{1}{2}}$ . Thus orthonormal Gram-Schmidt vectors are



$$z^{*(k)} = z^{(k)} Q^{(k)},$$

where  $Q^{(k)}$  is upper triangular with the reciprocals of the lengths of the Gram-Schmidt vectors in the diagonal.

Recapitulating at this point, we have an orthonormal basis for the active estimation space in terms of the Gram-Schmidt orthogonal vectors, where the last Gram-Schmidt vector was the component of the last estimation vector selected orthogonal to the space of the others.

It is interesting to note that the lengths of the Gram-Schmidt vectors  $z'_k$  are readily available from the original estimation vectors. In fact, using the basis  $z_1^*, \dots, z_k^*$  derived from the Gram-Schmidt vectors as the orthonormal basis of the previous section, it follows from the results of that section that  $z^{*(k)} = z^{(k)} Q^{(k)}$ , where  $Q^{(k)}$  is triangular with  $(z'_k \cdot z'_k)^{-\frac{1}{2}} = q_{kk}$ , and that  $h^{(k)-1} = Q^{(k)} Q^{(k)T}$ , or writing  $a^{(k)} = h^{(k)-1}$ , that  $a_{kk}^{(k)} = q_{kk}^2 = (z'_k \cdot z'_k)^{-1}$ .

Now, given orthonormal vectors,  $z_1^*, \dots, z_{k-1}^*, z_k^*$ , from the preceding section the square of the projection of  $y$  onto  $V_{k-1}$  was  $\sum_{i=1}^{k-1} b_i^{*2}$ , where

$$b^{*(k-1)} = z^{*(k-1)T} y;$$

while the square of the projection onto  $V_k$  is  $\sum_{i=1}^k b_i^{*2}$ , where

$$b^{*(k)} = z^{*(k)T} y.$$

Thus,  $b_k^{*2}$  is the increase in the square of the projection vector obtained by activating the estimation vector  $z_k$  (whose component orthogonal to  $V_{k-1}$  is  $z'_k$ ); or, equivalently,  $b_k^{*2}$  is the reduction in the square of the residual error vector obtained by activating  $z_k$ .

Now the principle of the step-up procedure becomes clear. Given the problem of augmenting by one vector an active estimation set of  $k-1$ , the answer is to choose that one for which the new projection of  $y$  in  $V_k$  has the largest component orthogonal to the old projection in  $V_{k-1}$ ; i.e., choose  $z_k$  so that relative to the augmented Gram-Schmidt orthonormal system,  $z_1^*, \dots, z_{k-1}^*, z_k^*, b_k^{*2}$  is maximum.

Again, it is important to be able to examine what values  $b_k^{*2}$  could have for the various possible vectors which could be chosen as  $z_k$ , and to do this easily in terms of the original vectors. But recall that

$$z^{*(k)} = z^{(k)} Q^{(k)}, \quad Q^{(k)} b^{*(k)} = b^{(k)} = h^{(k)-1} g^{(k)},$$

so that the triangularity of  $Q^{(k)}$  implies that

$$q_{kk} b_k^* = b_k^{(k)}, \quad \text{or } b_k^{*2} = \frac{b_k^{(k)2}}{a_{kk}^{(k)}}.$$

It is worth noting that the residual error vector can be considered as a final Gram-Schmidt vector, since  $e^{(k)} = y - \hat{y}^{(k)}$ , where  $\hat{y}^{(k)}$  is the projection of  $y$  onto  $V_k$ . But we have seen that the reciprocal of the square of the  $k$ th Gram-Schmidt vector is the last diagonal element of the inverse of  $h^{(k)}$ . Thus, if the  $h^{(k)}$  matrix being used is augmented with an additional column  $z^{(k)T} y$  and a symmetric row, corresponding to the dependent-variable vector  $y$ , then the last diagonal element of the inverse of this augmented matrix will be the reciprocal of the sum of least squares.

A computation synthesis of the procedure can be envisaged as a sequence of gaussian elimination tableaux, where starting with

$h_{11}$	...	$h_{1p}$	$g_1$	1	0	...	0	0
				0	$\ddots$	$\ddots$	$\vdots$	$\vdots$
$\vdots$		$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\ddots$	0	
$h_{p1}$	...	$h_{pp}$	$g_p$	0	...	0	1	0
$g_1$	...	$g_p$	$y^T y = G$	0	...	...	0	1

after  $k-1$  stages we have

1	0	...	0	$h_{1,k}^{(k-1)}$	...	$b_1^{(k-1)}$	$a_{11}^{(k-1)}$	...	$a_{1,k-1}^{(k-1)}$	0	...	0	0
0	$\ddots$		$\ddots$	$\vdots$		$\vdots$	$\vdots$		$\vdots$	$\vdots$		$\vdots$	$\vdots$
$\vdots$	$\ddots$		0										
0	...		0	1	$h_{k-1,k}^{(k-1)}$	...	$b_{k-1}^{(k-1)}$	$a_{k-1,1}^{(k-1)}$	...	$a_{k-1,k-1}^{(k-1)}$	0	...	0
0	...		...	0	$h_{kk}^{(k-1)}$	...	$g_k^{(k-1)}$	$\vdots$		$\vdots$	1	0	...
$\vdots$			$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$		$\vdots$	0	$\ddots$	$\vdots$
0	...		...	0			$g_p^{(k-1)}$				0	...	0
0	...		...	0	...	...	$G^{(k-1)}$	...	...	...	0	...	0

Note that  $\begin{bmatrix} h_{1k}^{(k-1)} \\ \vdots \\ h_{k-1,k}^{(k-1)} \end{bmatrix}$  is the solution of  $h^{(k-1)} x^{(k-1)} = v^{(k-1)}$  from

which the  $k$ th Gram-Schmidt vector is obtainable. Note that 
$$\begin{bmatrix} b_1^{(k-1)} \\ \vdots \\ b_{k-1}^{(k-1)} \end{bmatrix}$$
 is

the solution of  $h^{(k-1)} b^{(k-1)} = g^{(k-1)}$ . Note that if  $z_k$  is to be the next vector activated, then to obtain solutions to  $h^{(k)} x^{(k)} = v^{(k)}$  and  $h^{(k)} b^{(k)} = g^{(k)}$ , and to obtain  $a^{(k)} = h^{(k)-1}$ , requires only to operate on the above matrix with elementary (row) transformations so as to reduce the  $k$ th column to the unit vector  $u_k$ . This will produce

$$b_k^{(k)} = \frac{g_k^{(k-1)}}{h_{kk}^{(k-1)}} \quad \text{and} \quad a_{kk}^{(k)} = \frac{1}{h_{kk}^{(k-1)}}.$$

Thus  $b_k^{*2} = g_k^{(k-1)2} / h_{kk}^{(k-1)}$ .

From the last equation it is easy to see that, to find the vector yielding maximum  $b_k^{*2}$ , one need only examine the ratios  $(g_j^{(k-1)2}) / h_{jj}^{(k-1)}$  for  $j = k, k+1, \dots, p$ .

Note finally that, after  $k$  vectors have been chosen, the last diagonal element of the inverse of the augmented matrix would be  $1/G^{(k)}$ . Hence  $G^{(k)} = e^{(k)2}$ , the sum of squares of residual error.

Attention is called to the obvious fact that the step-up procedure of activating estimation vectors in the order of the further reduction made to sum of squares of error is not necessarily optimal in selecting say  $k$  vectors out of  $p$ . E.g. the  $y$  vector could be practically in the space of two vectors,  $z_1$  and  $z_2$ , but lying closer to a third  $z_3$  (not in the space) than to either of the given two. Thus the first vector selected would be vector  $z_3$ . Then regardless of which one was selected next, the pair chosen would be inferior to  $z_1, z_2$ .

One other word of caution is in order. The criterion for activating the next estimation vector is a maximum ratio. The denominator of this ratio is the square of the length of the component of the new vector in

the direction orthogonal to the then current estimation space. Of course, if some of the remaining vectors lie in the currently active estimation space (i.e., they are linearly dependent on vectors already chosen) they should not be considered as candidates. Because of roundoff errors such dependency must be defined approximately. Note that an almost dependent vector will produce a small orthogonal component which will tend to produce a large criterion ratio (which may be primarily an accident of roundoff error). To avoid spurious selections caused in this way the criterion should be compared only for those vectors whose orthogonal component exceeds a minimum value. What minimum value ought to be chosen is at this time a matter for conjecture.

4. Criterion for eliminating insignificant variables. From the discussion in the preceding section it evidently may happen that, in trying to activate an efficient set of  $k$  estimation vectors, the step-up procedure will select at one stage a vector which later on would be more efficiently eliminated. So far no procedure for deactivating any of the active estimation vectors has been incorporated. However, the algebraic technique for eliminating any designated active estimation vector and obtaining the regression analysis for the reduced set is well-known. It is a question of deciding whether to eliminate one and if so which one to eliminate. The purpose of this section is to provide a geometrically appealing and obvious answer to the second aspect of this question. Criteria for deciding whether to eliminate a variable will be discussed in the next section.

Therefore we suppose  $k$  estimation vectors have been activated and the corresponding analysis laid out, say in the manner of the sequence of gaussian tableaux referred to in the last section, and we suppose the decision has been made to eliminate one of the vectors. The question is: Which one shall we eliminate? Fix attention on one of the active  $z_1$ ,

say for definiteness the last one,  $z_k$ . Now the projection  $\hat{y}^{(k)}$  of  $y$  onto  $V_k$  can be resolved into its projection  $\hat{y}^{(k-1)}$  onto  $V_{k-1}$ , the space spanned by  $z_1, \dots, z_{k-1}$ , and a component orthogonal to  $\hat{y}^{(k-1)}$ . The projection  $\hat{y}^{(k-1)}$  of  $\hat{y}^{(k)}$  onto  $V_{k-1}$  is indeed the same as the direct projection of  $y$  onto  $V_{k-1}$ , so that the orthogonal component mentioned above in the resolution of  $\hat{y}^{(k)}$  is the net effect of the active vector  $z_k$  in the estimation of  $y$  with  $\hat{y}^{(k)}$ . Still keeping attention to  $z_k$ , we have already seen that the square of the length of this orthogonal component is  $b_k^{*2}$ , where in fact  $b_k^*$  is a component in the direction of the  $k$ th Gram-Schmidt vector generated according to the order in which the  $z_i$  were selected. Also,

$$b_k^{*2} = \frac{b_k^{(k)2}}{a_{kk}^{(k)}},$$

where, it will be recalled,

$$h^{(j)}_b(j) = g^{(j)}$$

for any  $j = 1, 2, \dots, p$ ; with  $h^{(j)} = z^{(j)T} z^{(j)}$ ,  $z^{(j)} = (z_1, \dots, z_j)$ ,  $a^{(j)} = h^{(j)-1}$ .

Recall also the pythagorean relation for each  $j = 1, 2, \dots, p$ ,

$$(y \cdot y) = y^2 = \hat{y}^{(j)2} + e^{(j)2},$$

where

$$\hat{y}^{(j)2} = \sum_{i=1}^j b_i^{*2} \text{ and } e^{(j)2} = \sum_{\mu=j+1}^N y_\mu'^2,$$

with  $y' = Py$ , the image of  $y$  under orthogonal transformation. Thus, remembering that  $b_i^* = y_i'$ ,

$$y^2 = y^{(k-1)2} + b_k^{*2} + e^{(k)2}.$$

Evidently  $b_k^{*2}$  can be interpreted as the net reduction in the square of the error vector obtained by activating  $z_k$ , or, equally as well, as the net increment (provided by activating  $z_k$ ) in the square of the active estimate.

Imagine now that the gaussian elimination has proceeded to the point of obtaining a solution to  $h^{(k)} b^{(k)} = g^{(k)}$  with  $a^{(k)} = h^{(k)-1}$ :

1	0 ... 0	$h_{1,k+1}^{(k)} \dots h_{1p}^{(k)}$	$b_1^{(k)}$	$a_{11}^{(k)} \dots$	0 ... 0	0
	$\vdots$	$\vdots$	$\vdots$	$\ddots$		
0	$\vdots$	$\vdots$	$b_j^{(k)}$	$a_{jj}^{(k)}$	$\vdots$	$\vdots$
	$\vdots$	$\vdots$	$\vdots$	$\ddots$		
0 ... 0	1	$h_{k,k+1}^{(k)} \dots h_{kp}^{(k)}$	$b_k^{(k)}$	$a_{kk}^{(k)}$	0 ... 0	0
<hr/>						
0	... 0	...	$g_{k+1}^{(k)}$	...	1 0 ... 0	0
	$\vdots$	$\vdots$	$\vdots$		$\ddots$	
	$\vdots$	$\vdots$	$\vdots$		0 $\ddots$ 0	$\vdots$
	$\vdots$	$\vdots$	$\vdots$		$\ddots$	
0 ... 0	0	...	$g_p^{(k)}$	...	0 ... 0 1	0
<hr/>						
0	... 0	...	$G^{(k)}$	...	0 ... 0	1

But now suppose  $j < k$  and the order in which  $z_j$  and  $z_k$  have been introduced

is reversed. Imagine re-scheduling the calculations in the gaussian elimination for this revision. In the tableaux this would be accomplished if in the initial tableau the  $j$ th and  $k$ th rows were interchanged and the  $j$ th and  $k$ th columns (to restore the initial unit matrix on the right the  $(p+1+j)$ -th and the  $(p+1+k)$ -th columns would also have to be interchanged), and thereafter repeating the operations which produced the  $k$ th tableau laid out above. The solution vector  $b^{(k)}$  in this case would be the same as before except that the order of  $b_j^{(k)}$  and  $b_k^{(k)}$  would be interchanged. Moreover, the inverse matrix would be the same except that the  $j$ th and  $k$ th rows and the  $j$ th and  $k$ th columns would be switched, putting  $a_{jj}^{(k)}$  in the  $(k,k)$ -position and  $a_{kk}^{(k)}$  in the  $(j,j)$ -position. Note now that  $b_j^{(k)2}/a_{jj}^{(k)}$  plays the role of  $b_k^{*2}$ , and hence the quantity  $b_j^{(k)2}/a_{jj}^{(k)}$  is the net reduction in the square of the error vector due to the  $z_j$  vector.

Now it is clear which of the  $k$  active estimation vectors should be eliminated, viz. that  $z_j$  ( $j \leq k$ ) for which  $b_j^{(k)2}/a_{jj}^{(k)}$  is minimum. Observe that these ratios are computable from the  $k$ th gaussian tableau set out above without any re-computations.

Having decided which estimation vector is to be eliminated from the active set of  $k$ , the procedure for making the elimination and obtaining the regression analysis for the reduced set of  $k-1$  active estimation vectors is as follows. According to the foregoing remarks no generatlity will be lost if we assume that the vector to be eliminated is  $z_k$ . But recall that to add  $z_k$  to the active set,  $z_1, \dots, z_{k-1}$ , and to obtain the regression analysis for the augmented set it was only necessary to perform

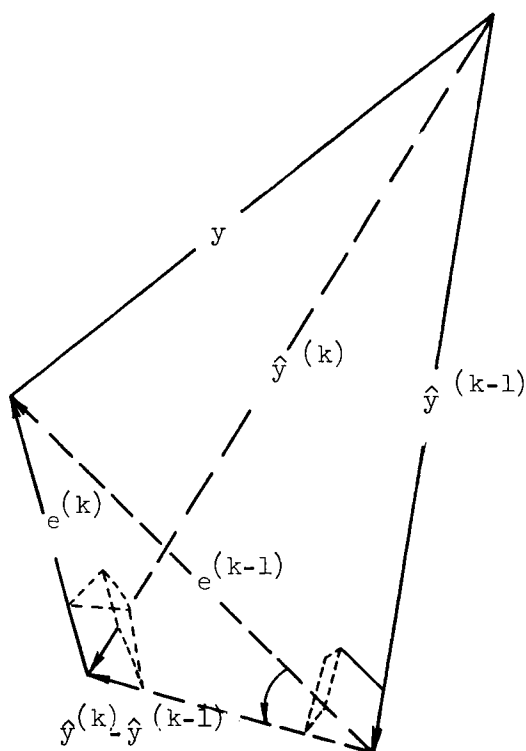


on the  $(k-1)$ -st tableau those elementary row transformations which reduce the  $k$ th column to the unit vector  $u_k$ . Therefore, to eliminate  $z_k$  it is only necessary to undo these calculations. It is not hard to verify that the reversing calculations are those elementary row transformation (on the  $k$ th tableau) which reduce the  $k$ th column of the inverse  $a^{(k)}$  back to  $u_k$ .

It is of course only a notational convenience to assume that the estimation vectors activated are the first  $k$  of the  $p$  listed in the tableaux. The swapping of rows and columns, while tidying up the written portrayal of the tableaux, etc., is completely unnecessary for computer handling of the problem.

Finally we shall mention that the rule described above for deciding which vector to eliminate is equivalent to that of eliminating the active vector that has the smallest partial correlation with the dependent variable vector. The partial correlation coefficient between  $z_k$  (say) and  $y$  is the cosine of the angle between  $e^{(k-1)}$  and  $\hat{y}^{(k)} - \hat{y}^{(k-1)}$ .

From the sketch below it is clear that this correlation decreases as the length  $||\hat{y}^{(k)} - \hat{y}^{(k-1)}|| = |b_k^*|$  decreases:



From the definition of cosine between  $e^{(k)}$  and  $\hat{y}^{(k)} - \hat{y}^{(k-1)}$  it is easy to show that

$$\cos^2 \theta(e^{(k-1)}, \hat{y}^{(k)} - \hat{y}^{(k-1)}) = \frac{b_k^{*2}}{(e^{(k-1)} \cdot e^{(k-1)})} = G^{(k-1)} b_k^{*2} = \frac{G^{(k-1)} b_k^{(k)2}}{a_{kk}^{(k)}}$$

5. Decision rules: the statistical model. In the last section the question answered was which active estimation variable ought to be eliminated once the decision had been made to eliminate one. The question of constructing decision rules to tell when to eliminate a variable was left for this section. Defining a sweep or iteration as a step in which either an inactive estimation vector is activated or an active one is deactivated, an obvious type of decision rule is the

following: Activate two vectors according to the step-up procedure, then eliminate one by the method described in the preceding section, and continue operating under this rule until some stopping rule (see below) stops the entire procedure. It is conceivable that such a rule would have utility if it is important in the ultimate application to have no more than  $k$  vectors while the cost of the extra sweeps is relatively unimportant.

Of course if of  $k$  active estimation vectors one has a partial correlation with the dependent variable vector of practically zero, it would seem wise to eliminate it. This suggests another quite arbitrary type of elimination rule: Of the  $k$  currently active estimation vectors eliminate the one of lowest partial correlation with  $y$  if said partial correlation is less than some level  $\alpha(k)$ , possibly a function of  $k$ .

Another decision problem must be dealt with, viz. that of constructing a stopping rule to stop the step-up procedure (with or without modification to allow for deletions). Here again, certain obvious but rather arbitrary rules come to mind. E.g., stop when  $k$  vectors have been activated (actually this was the somewhat naive rule used to motivate the section on the step-up procedure). It seems clear that, by itself, this is not a good rule, since in a particular example a satisfactory estimate may be attainable with far fewer than  $k$  vectors (i.e. the multiple correlation coefficient may be already very near one with fewer vectors or simply may not be improved "significantly" to warrant the inclusion of more).

We take the position at the present time of recommending a fairly comprehensive battery of stopping rules, any combination of which might be used, with a variety of sensitivity settings possible. Intuition suggests that appropriate settings will vary with the type of problem, the usage requirements and the burden of cost in time and money. Perhaps a battery of stopping rules should at least make provision for stopping when a fixed number of estimation vectors have been activated,

when the estimate is of sufficiently high accuracy (multiple correlation sufficiently near one), when the number of sweeps exceeds a certain number (this acts as a safeguard against a cyclic pattern of activation and elimination of vectors), and when the last  $r$  (say) vectors activated have not produced a "significant" change in the estimate.

Again the word, "significant", requires specific interpretation before the rule can be operational. One modus operandi might be: Stop the procedure if the increase in the multiple correlation coefficient  $R$ , produced by adding the last  $r$  active estimation vectors, was less than  $\beta(r,k)$ .

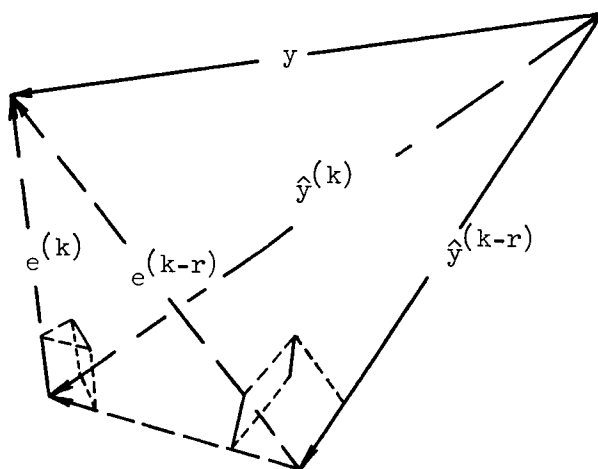
Both in the question of whether to deactivate an active estimation vector and in the question of when to stop activating estimation vectors the notion of significant effect arises. This suggests the possibility of resorting to a statistical model where the techniques of testing hypotheses might be invoked as a basis for decisions on whether to eliminate a variable or whether to stop the activation process.

In the remainder of this section we shall sketch the outline of a statistical model perhaps sufficiently to indicate the attractiveness of such a decision mechanism as well as to indicate some of the limitations of such a model.

Very briefly the model develops a statistic, or function of the observed active estimation vectors and the dependent variable vector, called an  $F$  statistic which is the decision-making instrument--large  $F$  means significance of the effects being tested and small  $F$  means nonsignificance. Under the hypothesis of the statistical model, and under the additional hypothesis that the effects of the estimation vectors being tested are only "noise" effects or effects introduced by virtue of random fluctuations, the  $F$  statistic is expected to have a value of about unity.

Actually, the  $F$  statistic is a ratio of the average of the effects of the vectors being tested to the average of some random error effects. In the terminology developed in previous sections suppose that

$z_{k-r+1}, \dots, z_k$  are active estimation vectors whose combined effect is being tested. Recall that  $\hat{y}^{(k)}$  is the projection of  $y$  on the space spanned by  $z_1, \dots, z_k$ ; and that  $\hat{y}^{(k-r)}$  is the projection of  $\hat{y}^{(k)}$  as well as the projection of  $y$  onto the subspace spanned by  $z_1, \dots, z_{k-r}$ .



In the F ratio the average of the effects of the  $r$  vectors  $z_{k-r+1}, \dots, z_k$  is measured as  $\frac{1}{r}$  times the square of the length of the vector,  $\hat{y}^{(k)} - \hat{y}^{(k-r)}$ ; while the average of error components is measured as  $\frac{1}{N-k}$  times the square of the so-called error vector,  $e^{(k)}$  (recall that  $e^{(k)}$  lies in a space of  $N-k$  dimensions orthogonal to the space generated by  $z_1, \dots, z_k$  in which  $\hat{y}^{(k)} - \hat{y}^{(k-r)}$  lies). Obviously, values of the F statistic less than one would not tend to support significant effects of  $z_{k-r+1}, \dots, z_k$ , while values greater than one presumably would. With the normal law of

errors assumed in the statistical model and under the hypothesis that these supposed effects of the last  $r$  vectors are noise effects, it turns out that the chances are approximately even that  $F$  should exceed the critical value of unity. If the critical value is increased the probability that the  $F$  statistic will exceed it diminishes rapidly. These probabilities are tabulated for various critical values and various degrees of freedom ( $r$  and  $N-k$  in our case). One may establish a decision rule to reject the hypothesis of no systematic effect (from the estimation vectors being tested) if the value of the  $F$  statistic observed is improbably larger than one.

The decision rule is not complete until specific numbers or functions are attached to the words "improbably larger." Undoubtedly a judicious choice depends on several factors involved in the balancing of cost and return in a particular problem. This is one of the open questions we have tried to study experimentally in another supporting study.

To complete the exposition some description of the characteristics of the assumed statistical model is warranted, although as we have mentioned there are recent excellent accounts of this model.

In the statistical linear regression model it is assumed that, except for random variations,  $Y$  is a linear function of the  $Z_i$ . Thus

$$y_u = \sum_{i=1}^p \beta_i z_{\mu i} + \epsilon_\mu, \mu = 1, 2, \dots, N,$$

where  $\epsilon_\mu$  are random errors. In addition it is usually assumed that the  $\epsilon_\mu$  are uncorrelated with a common variance  $\sigma^2$  and a mean of zero. The  $\beta_i$  are parameters which may be estimated in an optimal way under the

circumstances. In fact, the best linear unbiased estimate of a linear combination of the  $\beta_i$ , say  $\eta = \sum_{i=1}^p \beta_i Z_i$ , best in the sense of smallest variance, is  $\hat{Y} = \sum b_i Z_i$ , where the  $b_i$  are precisely those which produce the least squares estimate. This is the Gauss-Markov theorem.

It implies that, if the true functional relationship is except for a random error  $Y = \eta = \sum \beta_i Z_i$ , then, faced with not knowing the exact values of the  $\beta_i$ , the next best thing is to use the estimation function  $Y = \hat{Y} = \sum b_i Z_i$ .

To see the truth of this theorem we shall need to use the expected value or mean value operator  $E$  operating on a random variable or vector or matrix, with the expected value of a matrix of random variables being the matrix of expected values. From this definition it follows directly that  $E A X B = A(E X) B$ , if  $X$  is a random matrix and  $A$  and  $B$  are nonrandom matrices.

Now under the statistical model above,  $y = z\beta + \epsilon$ , where  $y^T = (y_1, \dots, y_N)$ ,  $z = (z_1, \dots, z_p)$ ,  $z_i^T = (z_{1i}, \dots, z_{Ni})$ ,  $\beta^T = (\beta_1, \dots, \beta_p)$ ,  $\epsilon^T = (\epsilon_1, \dots, \epsilon_N)$ , with  $\epsilon$  (and hence  $y$ ) being random vectors. According to the assumptions,  $E\epsilon = 0$  so that  $Ey = z\beta$ ; and the  $\epsilon_\mu$  are uncorrelated with a common variance  $\sigma^2$ , so that  $E\epsilon\epsilon^T = \sigma^2 I$ ,  $I$  being an identity matrix. Note that the  $z_i$  vectors are nonrandom.

First we show that  $Eb = \beta$ , i.e., that the  $b_i$  are unbiased estimates of the corresponding  $\beta_i$ . In fact

$$\begin{aligned} E\mathbf{b} &= E\mathbf{h}^{-1}\mathbf{g} = \mathbf{h}^{-1}E\mathbf{g} = \mathbf{h}^{-1}E\mathbf{z}^T\mathbf{y} = \mathbf{h}^{-1}\mathbf{z}^TE\mathbf{y} = \\ &\mathbf{h}^{-1}\mathbf{z}^T\mathbf{z}\boldsymbol{\beta} = \mathbf{h}^{-1}\mathbf{h}\boldsymbol{\beta} = \mathbf{I}\boldsymbol{\beta} = \boldsymbol{\beta}. \end{aligned}$$

Next we exhibit the covariance matrix of the estimates  $\mathbf{b}$ :

$$\begin{aligned} E(\mathbf{b} - E\mathbf{b})(\mathbf{b} - E\mathbf{b})^T &= E(\mathbf{b} - \boldsymbol{\beta})(\mathbf{b} - \boldsymbol{\beta})^T = \\ E(\mathbf{h}^{-1}\mathbf{g} - E\mathbf{h}^{-1}\mathbf{g})(\mathbf{h}^{-1}\mathbf{g} - E\mathbf{h}^{-1}\mathbf{g})^T &= \\ \mathbf{h}^{-1}E(\mathbf{g} - E\mathbf{g})(\mathbf{g} - E\mathbf{g})^T\mathbf{h}^{-1}, \end{aligned}$$

since  $\mathbf{h}$  and  $\mathbf{h}^{-1}$  are symmetric. Now

$$\begin{aligned} E(\mathbf{g} - E\mathbf{g})(\mathbf{g} - E\mathbf{g})^T &= E(\mathbf{z}^T\mathbf{y} - E\mathbf{z}^T\mathbf{y})(\mathbf{z}^T\mathbf{y} - E\mathbf{z}^T\mathbf{y})^T = \\ \mathbf{z}^TE(\mathbf{y} - E\mathbf{y})(\mathbf{y} - E\mathbf{y})^T\mathbf{z} &= \mathbf{z}^TE\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T\mathbf{z} = \mathbf{z}^T\sigma^2\mathbf{I}\mathbf{z} = \sigma^2\mathbf{h}. \end{aligned}$$

Hence, substituting above,

$$E(\mathbf{b} - E\mathbf{b})(\mathbf{b} - E\mathbf{b})^T = \mathbf{h}^{-1}\sigma^2\mathbf{h}\mathbf{h}^{-1} = \mathbf{h}^{-1}\sigma^2.$$

Now consider  $\hat{\mathbf{Y}} = \sum_i \mathbf{b}_i \mathbf{Z}_i = \mathbf{Z}^T\mathbf{b}$  as an estimate of  $\boldsymbol{\eta} = \sum \boldsymbol{\beta}_i \mathbf{Z}_i = \mathbf{Z}^T\boldsymbol{\beta}$ .

Observe that

$$\hat{\mathbf{Y}} = \mathbf{Z}^T\mathbf{b} = (\mathbf{Z}^T\mathbf{h}^{-1}\mathbf{z}^T)\mathbf{y} = \mathbf{a}^T\mathbf{y},$$

where  $\mathbf{a}^T = \mathbf{Z}^T\mathbf{h}^{-1}\mathbf{z}^T$ . This is what is meant by saying that  $\hat{\mathbf{Y}}$  is a linear estimate of  $\boldsymbol{\eta}$ ; i.e. it is a linear combination of the observed values of the random dependent variable  $\mathbf{Y}$ .

Also  $E\hat{\mathbf{Y}} = E\mathbf{Z}^T\mathbf{b} = \mathbf{Z}^TE\mathbf{b} = \mathbf{Z}^T\boldsymbol{\beta} = \boldsymbol{\eta}$ . Hence  $\hat{\mathbf{Y}}$  is an unbiased estimate of  $\boldsymbol{\eta}$ .



Finally we must show that the variance of  $\tilde{Y}$  is less than that of any other linear unbiased estimate of  $\eta$ . Suppose  $\tilde{Y}$  to be another linear unbiased estimate of  $\eta$ , so that  $Y = c_1 y_1 + \dots + c_N y_N = c^T y$ , and  $E c^T y = \eta$ .

Now consider vectors in euclidean N-space. Note that  $a = z(h^{-1}Z)$ , a vector lying in the estimation space spanned by the vectors  $z_1, \dots, z_p$ . We shall see that the vector  $a$  is the projection of  $c$  onto the space spanned by  $z_1, \dots, z_p$ . Since  $E a^T y = E c^T y$ , then  $0 = E(c-a)^T y = (c-a)^T E y = (c-a)^T z \beta$ . This identity can hold only if  $(c-a)^T z = 0$ . But this implies that

$$(c-a)^T a = (c-a)^T z(h^{-1}Z) = 0.$$

Hence  $a$  and  $c-a$  are orthogonal, and the pythagorean relation,  $c^2 = a^2 + (c-a)^2$ , holds.

The variance of  $\tilde{Y}$  is

$$\begin{aligned} E(\tilde{Y} - E\tilde{Y})^2 &= E(\tilde{Y} - E\tilde{Y})(\tilde{Y} - E\tilde{Y})^T \\ &= E(c^T y - E c^T y)(c^T y - E c^T y)^T \\ &= c^T E(y - E y)(y - E y)^T c = c^T E \epsilon \epsilon^T c \\ &= \sigma^2 c^T c = \sigma^2 \{a^T a + (c-a)^T (c-a)\} > \sigma^2 a^T a. \end{aligned}$$

But of course by the same reasoning the variance of  $\hat{Y}$  is  $\sigma^2 a^T a$ . This shows that  $\hat{Y}$  is of minimum variance.

To arrive at the F-statistic test for our decision rule in eliminating an estimation vector, or in stopping the activation of estimation vectors, additional assumptions are needed. Suppose that  $k$  of the estimation vectors,  $z_1, \dots, z_k$  has been activated, and it happens that  $Y = \sum_{i=1}^k \beta_i z_i + \epsilon$ , in short that the statistical model is valid with these  $k$  variables, so that

$y_{\mu} = \sum_{i=1}^k \beta_i z_{\mu i} + \epsilon_{\mu}$  or  $y = z^{(k)} \beta^{(k)} + \epsilon$ , when  $\epsilon^T = (\epsilon_1, \dots, \epsilon_N)$ . Suppose, in addition to the conditions that  $E\epsilon = 0$  and  $E\epsilon\epsilon^T = \sigma^2 I$ , we require that the  $\epsilon_{\mu}$  be normally distributed. Now suppose we wish to test the hypothesis ( $H_0$ ) that the last  $r$  parameters  $\beta_{k-r+1}, \dots, \beta_k$  are in fact all zero. (Accepting this hypothesis implies that the activation of the last  $r$  estimation variables adds nothing to the estimate available with the first  $k-r$  variables.)

The basic idea of such a test is to divide the sample space, i.e. the space of possible values of the vector  $y$ , into a rejection region  $R$  and its complement, an acceptance region, the ultimate decision rule being to reject  $H_0$  in case the observed value of  $y$  falls in  $R$ . Naturally, in order to make the test a discriminating or powerful one the points in the rejection region ought to be chosen roughly so as to maximize the probability of rejection when  $H_0$  is not true, while at the same time the probability of rejection when  $H_0$  is true should be kept below a certain bound. Such a test is approximately obtained by putting in  $R$  those points with highest "trade-off ratio," this ratio being essentially the ratio of the maximum of the probability density functions (pdf) over the entire family of pdf's defined by the admissible values of the parameters, to the maximum of the pdf's over the subfamily where the hypothesis  $H_0$  holds. This ratio is called the likelihood ratio  $\lambda$ . Such points of highest likelihood ratio are placed in  $R$  until the set is as large as it can be and still have the desired bound or the probability of rejection when  $H_0$  is true.

The optimal character of the likelihood ratio test for the problem at hand is given excellent treatment in SCHEFFÉ.

Let  $\Omega$  stand for the parameter space of admissible values of the parameters. In our case

$$\Omega = \{ \beta^{(k)}, \sigma^2 \mid -\infty < \beta^{(k)} < \infty, \sigma^2 > 0 \}.$$

Let  $\omega$  stand for the subset of  $\Omega$  where  $H_0$  is true; i.e.

$$\omega = \{ \beta^{(k)}, \sigma^2 \mid -\infty < \beta^{(k-r)} < \infty, \beta_{k-r+1} = \dots = \beta_k = 0, \sigma^2 > 0 \}.$$

According to the hypothesis of the model the  $\epsilon_\mu$  are normally distributed, uncorrelated (and hence independent) with common variance,  $\sigma^2$ . Thus the joint pdf of the random vector  $y$  is (for a parameter point in  $\Omega$ )

$$\begin{aligned} f(y; \beta^{(k)}, \sigma^2) &= \prod_{n=1}^N (2\pi\sigma^2)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2\sigma^2} (y_\mu - \sum_{i=1}^k \beta_i z_{\mu i})^2 \right\} \\ &= (2\pi\sigma^2)^{-N/2} \exp \left\{ -\frac{1}{2\sigma^2} (y - z^{(k)} \beta^{(k)})^T (y - z^{(k)} \beta^{(k)}) \right\}. \end{aligned}$$

Now to determine  $R$  it is necessary to maximize  $f$  over  $\Omega$  and over  $\omega$ , form the ratio  $\lambda$ , and select values of  $y$  for which this is highest.

$$R = \left\{ y \mid \lambda(y) = \frac{\sup_{\Omega} f}{\sup_{\omega} f} \geq \lambda_{\alpha} \right\},$$

where  $\lambda_{\alpha}$  is a critical value chosen so that

$$\Pr \{ y \in R \mid H_0 \text{ is true} \} \leq \alpha;$$

here  $\alpha$  is called the significance or rejection level of the test.

We recall now that a sum of squares of  $m$  normal independent random variables with mean zero and variance one ( $N(0,1)$ ) is said to be a Chi-square variable with  $m$  degrees of freedom. The ratio of the average of two such sums of squares of independent  $N(0,1)$  variables, with  $m_1$  terms in the numerator and  $m_2$  in the denominator, is called an F variable with  $m_1$  and  $m_2$  degrees of freedom. The probability distribution of the F variable is widely tabulated. The following result is the one pertinent to our problem. For a statistical linear regression model, where the errors are  $N(0, \sigma^2)$  independently distributed, the rejection region  $R$  of significance level  $\alpha$ , provided by the likelihood ratio criterion for rejecting  $H_0$  as described above, is given by

$$R = \left\{ y \mid \frac{(\hat{y}^{(k)} - \hat{y}^{(k-r)})^2 / r}{e^{(k)2} / (N-k)} \geq F_{r, N-k}^{(\alpha)} \right\},$$

where  $F_{r, N-k}^{(\alpha)}$  is the critical value in the  $F_{r, N-k}$  distribution for which  $\Pr \{ F_{r, N-k} \geq F_{r, N-k}^{(\alpha)} \} = \alpha$ .

The proof of this important theorem is obtained by constructing the likelihood ratio  $\lambda$ , in which the maximization problems are observed to be essentially the least squares problem, then reducing the inequality  $\lambda(y) \geq \lambda_\alpha$  which defines the rejection set to the form given in the conclusion. Used in the proof are: The orthogonal transform of  $y$  based on the Gram-Schmidt vectors  $z_1', \dots, z_{k-r}', z_{k-r+1}', \dots, z_k'$  and the fact that orthogonal transforms of normal vectors are normal. Although the proof is available in numerous references, we sketch it here.

Lemma 1. Let  $y$  be a vector of  $N(m_\mu, \sigma^2)$ , independent, random variables, and let  $y' = Py$  be an orthogonal transform of  $y$ . Then  $y'$  is a vector of  $N(m'_\mu, \sigma^2)$ , independent, random variables, with  $m' = \sum_{\mu}^N p_{\mu\nu} m_\nu$ , where  $P = (p_{\mu\nu})$ . Proof: Write  $m^T = (m_1, \dots, m_N)$ , and let  $G(\xi')$  be the distribution function of  $y'$ . Then

$$\begin{aligned} G(\xi') &= \Pr[y' \leq \xi'] = \Pr[Py \leq \xi'] = \Pr[\{y | Py \leq \xi'\}] \\ &= \int_{\{y | Py \leq \xi'\}} (2\pi\sigma^2)^{-N/2} \exp\left\{-\frac{1}{2\sigma^2} (y - m)^T (y - m)\right\} dy. \end{aligned}$$

Now, making the transformation  $y' = Py$  in the integral, the Jacobian of the transformation is the determinant of the orthogonal matrix  $P$ , hence in absolute value is one; the domain of integration is transformed into  $\{y' | y' \leq \xi'\}$ ; and the integrand becomes  $(2\pi\sigma^2)^{-N/2} \exp\left\{-\frac{1}{2\sigma^2} (y' - Pm)^T (y' - Pm)\right\}$ . Hence

$$G(\xi') = \prod_{\mu=1}^N \int_{-\infty}^{\xi'_\mu} (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2} (y'_\mu - m'_\mu)^2\right\} dy'_\mu,$$

so that obviously the  $y'_\mu$  are  $N(m'_\mu, \sigma^2)$ , independent.

It is a corollary of lemma 1 that, if  $\epsilon$  is a vector of  $N(0, \sigma^2)$ , independent variables and  $\epsilon' = P\epsilon$ ,  $P$  orthogonal, then  $\epsilon'$  is a vector of  $N(0, \sigma^2)$  independent variables.

Lemma 2. Let  $y = z^{(k)} \beta^{(k)} + \epsilon$  be a statistical linear regression model. Let  $z^{*(k)}$  be the matrix of orthonormal vectors generated from  $z_1, \dots, z_k$  by the Gram-Schmidt process, so that  $z^{*(k)} = z^{(k)} Q^{(k)}$  where  $Q^{(k)}$  is upper triangular. Let  $\beta^{*(k)} = Q^{(k)-1} \beta^{(k)}$ . Then  $\beta_{k-r+1} = \dots = \beta_k = 0$  if and only if  $\beta_{k-r+1}^* = \dots = \beta_k^* = 0$ .

Proof: Suppose  $\beta_{k-r+1} = \dots = \beta_k = 0$ ; it follows from the equation  $\beta^{*(k)} = Q^{(k)} = Q^{(k)-1}\beta$  and the fact that  $Q^{(k)-1}$  is upper triangular that  $\beta_k^* = 0$ , then  $\beta_{k-1}^* = 0$ , etc., until  $\beta_{k-r+1}^* = 0$ . The converse argument is the same.

Proof of the main theorem: By Lemma 2

$$\omega = \{\beta^{(k)}, \sigma^2 \mid -\infty < \beta^{*(k-r)} < \infty, \beta_{k-r+1}^* = \dots = \beta_k^* = 0, \sigma^2 > 0\},$$

and of course, since  $y = z^{(k)}\beta^{(k)} + \epsilon$  and

$$z^{*(k)}\beta^{*(k)} = z^{(k)}Q^{(k)}Q^{(k)-1}\beta^{(k)} = z^{(k)}\beta^{(k)}, \text{ then } y = z^{*(k)}\beta^{*(k)} + \epsilon.$$

Now

$$\lambda(y) = \frac{\sup\{f(y; \beta^{(k)}, \sigma^2) \mid (\beta^{(k)}, \sigma^2) \in \Omega\}}{\sup\{f(y; \beta^{(k)}, \sigma^2) \mid (\beta^{(k)}, \sigma^2) \in \omega\}}$$

$$\text{where } f = (2\pi\sigma^2)^{-N/2} \exp \left\{ -\frac{1}{2\sigma^2} \epsilon^T \epsilon \right\},$$

with  $\epsilon = y - z^{*(k)}\beta^{*(k)}$ . Clearly the extremizations in both cases can be obtained by first minimizing  $\epsilon^T \epsilon$  with respect to the  $\beta_i^*$ , substituting these back in, and maximizing the resulting expressions with respect to  $\sigma^2$ .

But minimizing  $\epsilon^T \epsilon$  is precisely the LS problem encountered before.

Using (as before) the orthogonal transform,  $y' = Py$  and  $\epsilon' = P\epsilon$  where the first  $k$  rows of  $P$  are  $z^{*(k)T}$ ,

$$\epsilon^T \epsilon = \epsilon'^T \epsilon' = (y'_1 - \beta_1^*)^2 + \dots + (y'_k - \beta_k^*)^2 + \sum_{\mu=k+1}^N y_\mu'^2.$$

Obviously  $\epsilon^T \epsilon$  over  $\Omega$  is minimized when  $\beta_i^* = y'_i$ ,  $i = 1, \dots, k$  with the value

of  $\epsilon^T \epsilon$  reducing to  $\sum_{\mu=k+1}^N y_\mu'^2 = e^{(k)2}$ ; while  $\epsilon^T \epsilon$  is minimized on  $\omega$  when

$\beta_i^* = y'_i$ ,  $i = 1, \dots, k-r$  (recall that  $\beta_{k-r+1}^* = \dots = \beta_k^* = 0$  in this case),

with the value of  $\epsilon^T \epsilon$  reducing to

$$\sum_{\mu=k-r+1}^N y_{\mu}'^2 = (\hat{y}^{(k)} - \hat{y}^{(k-r)})^2 + e^{(k)2}$$

in this case.

Substituting these extreme values back in and maximizing the numerator and denominator with respect to  $\sigma^2$ , gives for the numerator  $\hat{\sigma}^2 = \frac{e^{(k)2}}{N}$  and for the denominator  $\hat{\sigma}^2 = \frac{\hat{y}^{(k)} - \hat{y}^{(k-r)}^2 + e^{(k)2}}{N}$ .

Replacing these in the expression for  $\lambda(y)$  we get

$$\lambda(y) = \left[ \frac{\hat{\sigma}^2}{\hat{\sigma}^2} \right]^{N/2} = \left[ 1 + \frac{(\hat{y}^{(k)} - \hat{y}^{(k-r)})^2}{e^{(k)2}} \right]^{N/2}.$$

Now

$$\begin{aligned} R = \{y | \lambda(y) \geq \lambda_{\omega}\} &= \left\{y \left| \left[ 1 + \frac{(\hat{y}^{(k)} - \hat{y}^{(k-r)})^2}{e^{(k)2}} \right]^{N/2} \geq \lambda_{\omega} \right. \right\} \\ &= \left\{y \left| \frac{(\hat{y}^{(k)} - \hat{y}^{(k-r)})^2/r}{e^{(k)2}/(N-k)} \geq \frac{(\lambda_{\omega}^{2/N} - 1)(N-k)}{r} \right. \right\}. \end{aligned}$$

Finally, since  $(\hat{y}^{(k)} - \hat{y}^{(k-r)})^2 = \sum_{i=k-r+1}^k y_i'^2$  and  $e^{(k)2} = \sum_{\mu=k+1}^N y_{\mu}'^2$ ,

since by Lemma 1  $y' = Py$ , a vector of normal independent variables with common variance  $\sigma^2$ , and since under the hypothesis  $H_0$   $Ey_i' = 0$  ( $i=k-r+1, \dots, k$ ),

then the ratio

$$\frac{(\hat{y}^{(k)} - \hat{y}^{(k-r)})^2/r}{e^{(k)2}/(N-k)} = \frac{\sum_{i=k-r+1}^k (y_i'/\sigma)^2/r}{\sum_{\mu=k+1}^N (y_{\mu}'/\sigma)^2/(N-k)}$$

is a ratio of averages of sums of squares of  $N(0,1)$  independent random variables when  $H_0$  is true. That is, the likelihood ratio is equivalent to an F statistic when  $H_0$  is true.



# SELECTION OF SIGNIFICANT ESTIMATION VARIABLES IN A LEAST SQUARES PROBLEM: EMPIRICAL COMPUTER STUDIES

The object of these studies was to investigate the usefulness of the step-up procedure or modifications of it, in choosing a subset of a large number of estimation variables which is good in a least squares sense. In the first phase of these studies we wished to compare the step-up procedure with the procedure of finding the best subset at each stage. Because of the large number of matrix inversions required in the last method we could handle only a very small number of terms.

The results of the first phase are summarized in the two examples which follow. In the first run we note that the step-up procedure gave two terms with  $R^2 = 0.724$  whereas the best two terms give  $R^2 = 0.886$ .

## Phase One - Run 1

In this run the dependent variable was

$$F(X_1, X_2, X_3) = 3/(1+X_1^2+2X_3) .$$

The polynomial model was a balanced polynomial linear in  $X_1, X_2$ , and  $X_3$ , i.e.,  $a_1X_1+a_2X_2+a_3X_3+a_4X_1X_2+a_5X_1X_3+a_6X_2X_3+a_7X_1X_2X_3$ . The 125 data points were in a rectangular design with  $X_1 = .25(.25) 1.25$ ,  $X_2 = .25(.25) 1.25$ , and  $X_3 = .25(.25) 1.25$ . As will be noted in this run, the function  $F$  is actually independent of  $X_2$  and hence the estimation variables  $Z_2, Z_4, Z_6, Z_7$  should not enter the regression equation.

Step-up Procedure		Optimum Set	
Estimation Variables	$R^2$	Estimation Variables	$R^2$
5	.569815	5	.569815
5,3	.724129	3,1	.885715
5,3,1	.957606	3,1,5	.957606
5,3,1,2	.957615	3,2,1,5	.957615
5,3,1,2,4	.957631	3,2,1,5,4	.957631
5,3,1,2,4,6	.957632	3,2,6,1,5,4	.957632
5,3,1,2,4,6,7	.957634	3,2,6,1,5,4,7	.957634

Note that the step-up procedure did not select the optimum subset of two variables.

#### Phase One - Run 2

In this run the dependent variable and the polynomial model were the same as in Run 1. The 500 data points were in a rectangular design with  $X_1 = .25(.25)2.50$ ,  $X_2 = .25(.25) 2.50$ , and  $X_3 = .25(.25) 1.25$ .

Step-up Procedure		Optimal Set	
Estimation Variables	$R^2$	Estimation Variables	$R^2$
1	.702925	1	.702925
1,3	.884762	3,1	.884762
1,3,5	.963786	3,1,5	.963786
1,3,5,2	.963789	3,2,1,5	.963789
1,3,5,2,6	.963791	3,2,6,1,5	.963791
1,3,5,2,6,4	.963791	3,2,6,1,5,4	.963791
1,3,5,2,6,4,7	.963791	3,2,6,1,5,4,7	.963791

In this case, the step-up procedure gave the optimal subset in each case.

### Conclusions from Phase One

These runs indicated that some modification (e.g., a throw-out rule) might be helpful in obtaining a regression equation which would be close to the optimal. To investigate every possible regression equation even from a small set of terms is so time consuming that we did not use any example with a large number of terms in this phase.

### Phase Two

In this phase we used examples with a large number of terms. We used various throw-out criteria to investigate the relative merits of each. We did not find the optimal subsets.

### Summary of Phase Two

In the first 12 runs in this phase we used a balanced polynomial model to approximate the dependent variables

$$F_1(X_1, X_2, X_3) = (X_1^4 + X_2^3 + X_3^2) |X_1 + X_2 - \frac{\pi}{2} X_3|^{-1/2}$$

$$F_2(X_1, X_2, X_3) = \exp(-X_1^2 X_2 X_3)$$

$$F_3(X_1, X_2, X_3) = \sqrt{(X_1^2 + X_2^2 + X_3^2)}.$$

The results of these runs are tabulated below.

In the case of  $F_3 = \sqrt{X_1^2 + X_2^2 + X_3^2}$ , the 47-term polynomial fits very well with  $R^2 = 0.999972$ . In fact the 4 terms  $X_1 X_2$ ,  $X_2 X_3$ ,  $X_1^2$ ,  $X_2^2$  give a fit with  $R^2 = 0.962$ . The first 7 terms obtained by the stepwise procedure are  $X_1 X_2$ ,

$X_2X_3$ ,  $X_1^2$ ,  $X_2^2$ ,  $X_2^2X_3$ ,  $X_1$ ,  $X_3^2$ , and have  $R^2 = 0.992$ . With a throw-out criterion  $\geq 1.44$ , however, we find that  $X_1X_2$ ,  $X_1^2$ ,  $X_2^2$ ,  $X_2^2X_3$ ,  $X_1$ ,  $X_3$ ,  $X_2^2$  fit with  $R^2 = 0.996$ .

Now for the case  $F_2 = \exp(-X_1^2X_2X_3)$  we found that the 47-term polynomial fit with  $R^2 = 0.996$ . The first seven terms obtained by the step-up procedure were  $X_1$ ,  $X_2X_3$ ,  $X_1^2$ ,  $X_2^2$ ,  $X_3^2$ ,  $X_1^3X_2^2X_3^2$ ,  $X_1X_2X_3$ , and  $X_1^2X_2X_3$  with  $R^2 = 0.949$ . With a throw out criterion  $\geq 0.8$  we find that the seven terms  $X_2X_3$ ,  $X_1^2$ ,  $X_2^2X_3^2$ ,  $X_1^3X_2^2X_3^2$ ,  $X_1X_2X_3$ ,  $X_1^2X_2X_3$ , and  $X_1X_2^2X_3^2$  are a better seven and fit with  $R^2 = 0.965$ .

With a throw-out criterion  $\geq 4.9$  we find that the seven terms  $X_1$ ,  $X_2X_3$ ,  $X_1^2X_3^2$ ,  $X_1X_2X_3$ ,  $X_1^2X_2X_3$ ,  $X_1X_2^2X_3^2$ ,  $X_1^3X_2^3X_3^2$  fit with  $R^2 = 0.962$  and that the seven terms  $X_1$ ,  $X_1X_2X_3$ ,  $X_1^2X_2X_3$ ,  $X_1X_2^2X_3^2$ ,  $X_1^3X_2^3X_3^2$ ,  $X_1^2X_2^2X_3$ , and  $X_1^2X_2$  fit with  $R^2 = 0.978$ . We also find in fact that the first five terms in the last fit have  $R^2 = 0.962$ . Thus the five terms  $X_1$ ,  $X_1X_2X_3$ ,  $X_1^2X_2X_3$ ,  $X_1X_2^2X_3^2$ , and  $X_1^3X_2^3X_3^2$  fit better than the seven terms given by the step-up procedure with no throw-out criterion.

$$\text{In case } F_1 = \frac{X_1^4 + X_2^3 + X_3^2}{\sqrt{|X_1 + X_2 - \frac{\pi}{2}X_3|}} \text{ where the denominator has zeros in the}$$

region of fitting we find that the fit is not quite as good. The 47-term polynomial gives  $R^2 = 0.938$ . Again, however, we find that a seven-term polynomial will do almost as well. The straight step-up procedure gives the seven terms  $X_1^3X_3$ ,  $X_1^3$ ,  $X_3^2$ ,  $X_2^3$ ,  $X_1^2X_2X_3$ ,  $X_1^3X_2^2X_3$ , and  $X_2X_3$  which fit with  $R^2 = 0.894$ . With a throw-out criterion  $\geq 6.3$  we find that the

seven terms  $X_1^3$ ,  $X_2^3$ ,  $X_1^2X_3^2$ ,  $X_1X_2X_3^2$ ,  $X_2X_3^2$ ,  $X_1X_2^2X_3^2$ , and  $X_2^3X_3^2$  fit with  $R^2 = 0.902$ .

This example also gave rise to the situation where, while  $X_1^3X_3$  is the best single term, it is not one of the best two terms. The best two terms involving  $X_1^3X_3$  are  $X_1^3X_3$  and  $X_1^3$  which fit with  $R^2 = 0.733$ . However, the two terms  $X_1^3$  and  $X_3^2$  fit with  $R^2 = 0.775$ . Another situation which occurred on this example was that with a throw-out criterion of  $\geq 4.9$  we would arrive at a five-term polynomial with  $R^2 = 0.876$  whereas the step-up procedure with no throw-out criterion leads to a five-term polynomial with  $R^2 = 0.884$ . Hence, having a throw-out criterion is not always better.

As an example of a non-balanced design with an arbitrary linear model we used a correlation matrix given in Anderson and Fruchter, "Prediction Selection Method," Psychometrika, Vol. 25, No. 1. The results are tabulated in Run 17. Here we found that the throw-out criterion was not used, and so the variables were selected by the step-up procedure without this option. The overall fit using 14 variables gave  $R^2 = 0.270$  and an  $F(14, 295) = 7.8$  which is significant at 0.005. However, an F test of the hypothesis that the last 9 variables have zero coefficients is not significant at even the 50% level. The  $R^2$  for the first five terms of the step-up procedure is  $R^2 = 0.259$ .

#### Phase Two - Run 1

In this run, the dependent variable was  $F(X_1, X_2, X_3) = (X_1^4 + X_2^3 + X_3^2) |X_1 + X_2 - \frac{\pi}{2}X_3|^{-1/2}$ . To fit this expression we used the polynomial

model

$$\sum_{l_1=0}^3 \sum_{l_2=0}^3 \sum_{l_3=0}^2 a_{l_1 l_2 l_3} X_1^{l_1} X_2^{l_2} X_3^{l_3}.$$

All the terms, including the dependent variable are first adjusted for their means. Thus we wish to find subsets of the 47 terms in this polynomial which give the best approximation to the dependent variable. The values of  $F(X_1, X_2, X_3)$  and  $X_1^{l_1} X_2^{l_2} X_3^{l_3}$  were all calculated at 500 points in a balanced design. In this run we used the points  $X_1 = .25(.25) 2.50$ ,  $X_2 = .25(.25) 2.50$ , and  $X_3 = .25(.25) 1.25$ .

The throw-out criterion for this run was  $F_0 = 1.5$ . A tabulation of the terms as they were brought in follows. (Reduced  $R^2$  is  $1 - \frac{N-1}{N-m} (1-R^2)$  where  $R^2$  is the square of the multiple correlation coefficient and  $N = 500$ , the number of observations, and  $m$  is the number of terms in the model.)

Sweep	<sup>m</sup> Terms in Model	Term No	Term	F in	F out	$R^2$	Reduced $R^2$
1	1	37	$X_1^3 X_3$	1058		.680	.680
2	2	36	$X_1^3$	98.96		.733	.733
3	3	2	$X_3^2$	99.34		.778	.777
4	4	9	$X_2^3$	103.15		.816	.815
5	5	28	$X_1^2 X_2 X_3$	292.59		.884	.884
6	6	43	$X_1^3 X_2^2 X_3$	25.37		.890	.889
7	7	4	$X_2 X_3$	19.01		.894	.893
8	8	14	$X_1 X_3^2$	26.76		.900	.898

<u>Sweep</u>	<u>m</u> <u>Terms</u> <u>in Model</u>	<u>Term No</u>	<u>Term</u>	<u>F in</u>	<u>F out</u>	<u>R<sup>2</sup></u>	<u>Reduced</u> <u>R<sup>2</sup></u>
9	9	16	$X_1 X_2 X_3$	15.58		.903	.901
10	8	2	$X_3^2$		0.21	.903	.901
11	9	17	$X_1 X_2 X_3^2$	8.34		.904	.903
12	10	19	$X_1 X_2^2 X_3$	10.78		.906	.905
13	9	43	$X_1^3 X_2^2 X_3$		0.11	.906	.905
14	8	37	$X_1^3 X_3$		1.47	.906	.905
15	9	1	$X_3$	12.94		.909	.907
16	10	10	$X_2^3 X_3$	11.01		.911	.909
17	11	45	$X_1^3 X_2^3$	15.92		.913	.912
18	12	5	$X_2 X_3^2$	14.62		.916	.914
19	13	38	$X_1^3 X_3^2$	6.96		.917	.915
20	14	2	$X_3^2$	6.66		.918	.916
21	13	1	$X_3$		0.15	.918	.916
22	14	40	$X_1^3 X_2 X_3$	6.69		.919	.917
23	13	28	$X_1^2 X_2 X_3$		0.04	.919	.917
24	14	43	$X_1^3 X_2^2 X_3$	3.23		.920	.918
25	15	44	$X_1^3 X_2^2 X_3^2$	2.84		.920	.918
26	16	12	$X_1$	3.93		.921	.918
27	17	11	$X_2^3 X_3^2$	3.27		.921	.919
28	16	4	$X_2 X_3$		0.02	.921	.919
29	17	21	$X_1 X_2^3$	1.62		.922	.919

## Run 2

In this run the dependent variable, the polynomial model and the data points were all the same as in Run 1. The throw-out criterion was  $F_0 = 0.9$ . This run should tend to throw out terms less often than Run 1. This should lead to fewer sweeps to reach k terms but perhaps the fit for these terms will not be as good as in Run 1. The tabulation through Sweep 13 is the identical with Run 1.

<u>Sweep</u>	<u>m</u> <u>Terms</u> <u>in Model</u>	<u>Term No</u>	<u>Term</u>	<u>F in</u>	<u>F out</u>	<u>R<sup>2</sup></u>	<u>Reduced</u> <u>R<sup>2</sup></u>
13	9	43	$x_1^3 x_2^2 x_3$		0.11	.906	.905
14	10	1	$x_3$	11.46		.909	.907
15	9	37	$x_1^3 x_3$		0.05	.909	.907
16	10	10	$x_2^3 x_3$	11.01		.911	.909

Sweeps 16 through 29 are the same as Run 1

29	17	21	$x_1 x_2^3$	1.62		.922	.919
30	16	45	$x_1^3 x_2^3$		0.03	.922	.919
31	17	26	$x_1^2 x_2^2 x_3$	1.34		.922	.919
32	18	24	$x_1^2$	3.09		.922	.920
33	19	13	$x_1 x_2$	1.20		.923	.920
34	20	18	$x_1 x_2^2$	2.86		.923	.920
35	19	21	$x_1 x_2^3$		0.00	.923	.920
36	20	1	$x_3$	1.01		.923	.920
37	19	11	$x_2^3 x_3$		0.59	.923	.920



<u>Sweep</u>	<u>Terms in Model</u>	<u>Term No</u>	<u>Term</u>	<u>F in</u>	<u>F out</u>	<u>R<sup>2</sup></u>	<u>Reduced R<sup>2</sup></u>
38	20	37	$x_1^3 x_3$	1.42		.923	.920
39	21	47	$x_1^3 x_2^3 x_3^2$	1.53		.924	.920
40	22	45	$x_1^3 x_2^3$	2.13		.924	.921
41	23	7	$x_2^2 x_3$	1.94		.924	.921
42	22	16	$x_1 x_2 x_3$		0.31	.924	.921
43	23	23	$x_1 x_2^3 x_3^2$	1.10		.924	.921
44	24	11	$x_2^3 x_3^2$	1.22		.925	.921
45	25	25	$x_1^2 x_3$	0.92		.925	.921

### Run 3

In this run the dependent variable, the polynomial model and the data points were all the same as in Run 1. The throw-out criterion for Run 3 was  $F_0 = 8.0$ . This run should tend to throw out terms more often than Run 1 or Run 2. This should lead to more sweeps to reach k terms but hopefully the fit for these k terms will be better than in Run 1 or Run 2. (Compare, however, Run 3, Sweep 7, with Run 1, Sweep 5 and also Run 3, Sweep 18 with Run 1, Sweep 12). Note that in Run 3 we see that the best term No. 37 is not one of the best two terms.

<u>Sweep</u>	<u>Terms in Model</u>	<u>Term No</u>	<u>Term</u>	<u>F in</u>	<u>F out</u>	<u>R<sup>2</sup></u>	<u>Reduced R<sup>2</sup></u>
1	1	37	$x_1^3 x_3$	1058.24		.680	.680
2	2	36	$x_1^3$	98.96		.733	.733
3	3	2	$x_3^2$	99.34		.778	.777
4	2	37	$x_1^3 x_3$		4.88	.775	.775

<u>Sweep</u>	<u>m</u> <u>Terms</u> <u>in Model</u>	<u>Term No</u>	<u>Term</u>	<u>F in</u>	<u>F out</u>	<u>R<sup>2</sup></u>	<u>Reduced</u> <u>R<sup>2</sup></u>
5	3	9	$x_2^3$	102.13		.814	.813
6	4	15	$x_1 x_2$	219.83		.871	.870
7	5	20	$x_1 x_2^2 x_3^2$	18.48		.875	.875
8	6	26	$x_1^2 x_3^2$	40.36		.885	.884
9	7	17	$x_1 x_2 x_3^2$	22.30		.890	.889
10	6	20	$x_1 x_2^2 x_3^2$		5.12	.889	.888
11	7	5	$x_2 x_3^2$	40.48		.897	.896
12	8	20	$x_1 x_2^2 x_3^2$	20.02		.901	.900
13	7	15	$x_1 x_2$		6.22	.900	.899
14	8	11	$x_2^3 x_3^2$	12.04		.903	.901
15	7	2	$x_3^2$		2.93	.902	.901
16	8	38	$x_1^3 x_3^2$	26.81		.907	.906
17	9	18	$x_1 x_2^2$	22.83		.911	.910
18	10	41	$x_1^3 x_2 x_3^2$	8.20		.912	.911

#### Run 4

In this run the dependent variable, the polynomial model and the data points were all the same as in Run 1. The throw-out criterion was  $F_0 = 10^{-3}$ . This run should not throw out variables very often, at least not until they are very insignificant. A partial tabulation of this run follows.

<u>Sweep</u>	<u>m</u> <u>Terms</u> <u>in Model</u>	<u>Term No</u>	<u>Term</u>	<u>F in</u>	<u>F out</u>	<u>R<sup>2</sup></u>	<u>Reduced</u> <u>R<sup>2</sup></u>
1	1	37	$x_1^3 x_3$	1058.00		.680	.680
2	2	36	$x_1^3$	98.96		.733	.733
3	3	2	$x_3^2$	99.34		.778	.777
4	4	9	$x_2^3$	103.15		.816	.815
5	5	28	$x_1^2 x_2 x_3$	292.59		.884	.884
6	6	43	$x_1^3 x_2^2 x_3$	25.37		.890	.889
7	7	4	$x_2 x_3$	19.01		.894	.893
8	8	14	$x_1 x_3^2$	26.76		.900	.898
9	9	16	$x_1 x_2 x_3$	15.58		.903	.901
10	10	41	$x_1^3 x_2 x_3^2$	10.40		.905	.903
11	11	45	$x_1^3 x_2^3$	11.46		.907	.905
12	12	21	$x_1 x_2^3$	23.71		.911	.909
13	13	46	$x_1^3 x_2^3 x_3$	6.60		.913	.910
14	12	9	$x_2^3$		0.00	.913	.911
15	13	1	$x_3$	3.93		.913	.911
16	14	10	$x_2^3 x_3$	3.10		.914	.911
20	16	2	$x_3^2$		0.00	.916	.914
25	21	25	$x_1^2 x_3$	3.67		.920	.917
30	22	24	$x_1^2$		0.00	.923	.920
35	27	39	$x_1^3 x_2$	5.35		.927	.923
40	30	9	$x_2^3$		0.00	.92828	.92385
45	31	20	$x_1 x_2^2 x_3^2$	0.53		.92866	.92410
50	34	34	$x_1^2 x_2^3 x_3$	15.23		.93196	.92714

<u>Sweep</u>	<u>m</u> <u>Terms</u> <u>in Model</u>	<u>Term No</u>	<u>Term</u>	<u>F in</u>	<u>F out</u>	<u>R<sup>2</sup></u>	<u>Reduced</u> <u>R<sup>2</sup></u>
55	37	6	$x_2^2$		0.00	.93459	.92950
60	40	42	$x_1^3 x_2^2$	0.44		.93682	.93146
65	45	18	$x_1 x_2^2$	0.55		.93730	.93124
66	46	9	$x_2^3$	1.04		.93745	.93125
67	47	30	$x_1^2 x_2^2$	5.08		.938142	.931860

#### Run 5

In this run, the dependent variable was  $F(X_1, X_2, X_3) = \exp(-X_1^2 X_2 X_3)$ . We used the same balanced polynomial model as in the first four runs, cubic in  $X_1$  and  $X_2$ , quadratic in  $X_3$ . The 500 data points were in the same balanced design,  $X_1 = .25(.25) 2.50$ ,  $X_2 = .25(.25) 2.50$ ,  $X_3 = .25(.25) 1.25$ . The polynomial model in this case should fit better than in the first four runs.

The throw-out criterion in the first runs in this series was  $F_0 = 1.5$ .

<u>Sweep</u>	<u>m</u> <u>Terms</u> <u>in Model</u>	<u>Term No.</u>	<u>Term</u>	<u>F in</u>	<u>F out</u>	<u>R<sup>2</sup></u>	<u>Reduced</u> <u>R<sup>2</sup></u>
1	1	12	$x_1$	836.43		.627	.627
2	2	4	$x_2 x_3$	529.77		.819	.819
3	3	24	$x_1^2$	212.65		.874	.873
4	4	8	$x_2^2 x_3^2$	308.40		.922	.922
5	5	44	$x_1^3 x_2^2 x_3^2$	61.94		.931	.930
6	6	16	$x_1 x_2 x_3$	103.35		.943	.942
7	7	28	$x_1^2 x_2 x_3$	62.76		.949	.949
8	8	20	$x_1 x_2^2 x_3^2$	231.18		.965	.965

<u>Sweep</u>	<u>m</u> <u>Terms</u> <u>in Model</u>	<u>Term No</u>	<u>Term</u>	<u>F in</u>	<u>F out</u>	<u>R<sup>2</sup></u>	<u>Reduced</u> <u>R<sup>2</sup></u>
9	7	12	$X_1$		0.79	.965	.965
10	8	23	$X_1 X_2^3 X_3^2$	27.03		.967	.967
11	9	22	$X_1 X_2^3 X_3$	138.53		.974	.974
12	10	21	$X_1 X_2^3$	411.57		.986	.986
13	11	27	$X_1^2 X_2$	8.79		.986	.986
14	12	25	$X_1^2 X_3$	97.48		.989	.988
15	13	30	$X_1^2 X_2^2$	67.45		.990	.990
16	14	38	$X_1^3 X_2^2$	73.37		.991	.991
17	13	24	$X_1^2$		0.03	.991	.991
18	14	33	$X_1^2 X_2^3$	32.08		.992	.992
19	15	39	$X_1^3 X_2$	23.02		.992	.992
20	16	32	$X_1^2 X_2^2 X_3^2$	75.84		.993	.993
25	17	42	$X_1^3 X_2^2$	33.90		.99431	.99412
26	18	35	$X_1^2 X_2^3 X_3^2$	14.09		.99449	.99428
27	17	23	$X_1 X_2^3 X_3^2$		1.20	.99446	.99427
28	18	8	$X_2^2 X_3^2$	11.81		.99459	.99440
29	19	14	$X_1 X_3^2$	18.00		.99479	.99459
30	20	45	$X_1^3 X_2^3$	5.57		.99485	.99464
35	25	21	$X_1 X_2^3$	11.63		.99594	.99573
40	26	21	$X_1 X_2^3$		1.12	.99604	.99583
44	26	21	$X_1 X_2^3$	0.93		.99606	.99585

### Run 6

This run used the same dependent variable, polynomial model and data points as in Run 5. The throw-out criterion was  $F_0 = 0.9$ . This will tend to throw out terms less often than in Run 5. In fact, however, the runs are identical through Sweep 26.

<u>Sweep</u>	<u>m</u> <u>Terms</u> <u>in Model</u>	<u>Term No</u>	<u>Term</u>	<u>F in</u>	<u>F out</u>	<u>R<sup>2</sup></u>	<u>Reduced</u> <u>R<sup>2</sup></u>
25	17	42	$x_1^3 x_2^2$	33.90		.99431	.99412
26	18	35	$x_1^2 x_2^3 x_3^2$	14.09		.99447	.99428
27	19	11	$x_2^3 x_3^2$	13.66		.99462	.99492
28	18	20	$x_1 x_2^2 x_3^2$		0.00	.99462	.99443
29	19	14	$x_1 x_3^2$	16.67		.99480	.99461
30	20	45	$x_1^3 x_2^3$	6.19		.99487	.99467
35	23	42	$x_1^3 x_2^2$		0.44	.99574	.99555
40	26	11	$x_2^3 x_3^2$		0.48	.99609	.99588
41	25	38	$x_1^3 x_3^2$		0.59	.99608	.99589
42	24	14	$x_1 x_3^2$		0.61	.99608	.99589
43	25	31	$x_1^2 x_2^2 x_3$	0.79		.99608	.99589

### Run 7

In this run the dependent variable, the polynomial model and the data points were all the same as in Run 5. The throw-out criterion was  $F_0 = 8.0$ . The variables brought in were the same as in Run 5 through Sweep 7.

<u>Sweep</u>	<u>m</u> <u>Terms</u> <u>in Model</u>	<u>Term No</u>	<u>Term</u>	<u>F in</u>	<u>F out</u>	<u>R<sup>2</sup></u>	<u>Reduced</u> <u>R<sup>2</sup></u>
7	7	28	$X_1^2 X_2 X_3$	62.76		.94925	.94863
8	6	24	$X_1^2$		4.41	.94879	.94827
9	5	44	$X_1^3 X_2^2 X_3^2$		4.83	.94829	.94787
10	6	20	$X_1 X_2^2 X_3^2$	69.96		.95471	.95425
11	7	47	$X_1^3 X_2^3 X_3^2$	100.42		.96239	.96193
12	6	4	$X_2 X_3$		1.81	.96225	.96187
13	5	8	$X_2^2 X_3^2$		1.07	.96217	.96186
14	6	31	$X_1^2 X_2^2 X_3$	63.95		.96651	.96617
15	7	27	$X_1^2 X_2$	269.22		.97836	.97809
16	8	23	$X_1 X_2^3 X_3^2$	123.13		.98270	.98245
17	9	4	$X_2 X_3$	73.73		.98496	.98471
18	10	25	$X_1^2 X_3$	72.64		.98690	.98666
19	11	26	$X_1^2 X_3^2$	50.82		.98814	.98790
20	12	36	$X_1^3$	52.25		.98929	.98905
21	13	44	$X_1^3 X_2^2 X_3^2$	46.61		.99023	.98999
22	14	30	$X_1^2 X_2^2$	47.15		.99109	.99085
23	15	33	$X_1^2 X_2^3$	78.01		.99233	.99211
24	14	47	$X_1^3 X_2^3 X_3^2$		0.88	.99232	.99211
25	13	12	$X_1$		3.00	.99227	.99208
26	14	8	$X_2^2 X_3^2$	55.18		.99306	.99287
27	15	22	$X_1 X_2^3 X_3$	6.18		.99314	.99295

# Run 8

In this run the dependent variable, the polynomial model and the data points were all the same as in Run 5. The throw-out criterion was  $F_0 = 10^{-3}$ . The variables brought in were the same as in Run 5 through Sweep 8.

Sweep	m Terms in Model	Term No	Term	F in	F out	$R^2$	Reduced $R^2$
8	8	20	$X_1 X_2^2 X_3^2$	251.18		.96550	.96501
9	9	23	$X_1 X_2^3 X_3^2$	26.72		.96728	.96675
10	10	22	$X_1 X_2^3 X_3$	137.60		.97447	.97400
11	11	21	$X_1 X_2^3$	416.70		.98623	.98595
12	12	36	$X_1^3$	32.83		.98710	.98681
13	13	40	$X_1^3 X_2 X_3$	50.05		.98830	.98801
14	14	15	$X_1 X_2$	9.58		.98853	.98822
15	15	13	$X_1 X_3$	148.72		.99122	.99097
20	20	6	$X_2^2$	20.37		.99481	.99461
25	25	33	$X_1^2 X_2^3$	10.32		.95537	.99514
30	30	7	$X_2^2 X_3$	4.12		.99576	.99550
35	35	30	$X_1^2 X_2^2$	3.22		.99619	.99591
40	38	17	$X_1 X_2 X_3^2$	5.45		.99630	.99601
45	43	38	$X_1^3 X_3^2$	2.68		.99638	.99605
50	44	8	$X_2^2 X_3^2$	0.87		.99640	.99606
54	46	32	$X_1^2 X_2^2 X_3^2$	0.18		.996399	.996042
55	47	46	$X_1^3 X_2^3 X_3$	2.21		.996416	.996052
56	46	12	$X_1$		0.00	.996416	.996061
57	47	12	$X_1$	0.00		.996416	.996052



# Run 9

In this run the dependent variable was  $F(X_1, X_2, X_3) = \sqrt{X_1^2 + X_2^2 + X_3^2}$ .

The 47-term balanced polynomial, cubic in  $X_1$  and  $X_2$  and quadratic in  $X_3$ , was used as the model to fit the dependent variable over the 500 data points  $X_1 = .25(.25) 2.50$ ,  $X_2 = .25(.25) 2.50$ , and  $X_3 = .25(.25) 1.25$ .

As expected in this case, the fit is very good. Because of the symmetry involved the terms in  $X_1$  and  $X_2$  should be the same, at least in the complete model. The lack of symmetry in the way these terms were brought is interesting.

The throw-out criterion for this run was  $F_0 = 1.5$ .

<u>Sweep</u>	<u>m</u> <u>Terms</u> <u>in Model</u>	<u>Term No</u>	<u>Term</u>	<u>F in</u>	<u>F out</u>	<u>R<sup>2</sup></u>	<u>Reduced</u> <u>R<sup>2</sup></u>
1	1	15	$X_1 X_2$	1337.01		.72861	.72861
2	2	4	$X_2 X_3$	98.19		.77338	.77293
3	3	24	$X_1^2$	309.02		.86037	.85981
4	4	6	$X_2^2$	1324.25		.96201	.96178
5	5	7	$X_2^2 X_3$	302.54		.97644	.97625
6	6	12	$X_1$	553.58		.98890	.98879
7	7	2	$X_3^2$	202.14		.99213	.99204
8	8	3	$X_2$	427.10		.99579	.99573
9	9	4	$X_2 X_3$		1.44	.99578	.99573
10	8	14	$X_1 X_3^2$	469.05		.99784	.99781
11	9	19	$X_1 X_2^2 X_3$	169.84		.99840	.99837
12	10	5	$X_2 X_3^2$	177.58		.99882	

<u>Sweep</u>	<u>m</u> <u>Terms</u> <u>in Model</u>	<u>Term No</u>	<u>Term</u>	<u>F in</u>	<u>F out</u>	<u>R<sup>2</sup></u>	<u>Reduced</u> <u>R<sup>2</sup></u>
13	11	9	$x_2^3$	144.20		.99909	
14	12	36	$x_1^3$	204.27		.99936	
15	13	17	$x_1 x_2 x_3^2$	171.96		.99953	
16	14	21	$x_1 x_2^3$	87.04		.99960	
17	15	39	$x_1^3 x_2$	59.18		.99964	
18	16	30	$x_1^2 x_2^2$	125.68		.99972	
19	17	1	$x_3$	55.81		.99975	
20	18	4	$x_2 x_3$	114.22		.99980	
25	23	16	$x_2 x_1 x_3$	107.82		.99994	
30	26	38	$x_1^3 x_3^2$	40.36		.99996	
35	27	20	$x_1 x_2^2 x_3^2$		0.09	.99996	
40	30	21	$x_1 x_2^3$		0.00	.99996	
45	35	45	$x_1^3 x_2^3$	9.43		.999969	
46	34	39	$x_1^3 x_2$		0.26	.999969	
47	35	21	$x_1 x_2^3$	3.23		.999969	
48	36	41	$x_1^3 x_2 x_3^2$	0.40		.999969	

#### Run 10

In this run the dependent variable, the polynomial model and the data points were the same as in Run 9. The throw-out criterion for this run was  $F_0 = 0.9$ . The tabulation of the results is identical with Run 9 through Sweep 8.

<u>Sweep</u>	<u>m</u> <u>Terms</u> <u>in Model</u>	<u>Term No</u>	<u>Term</u>	<u>F in</u>	<u>F out</u>	<u>R<sup>2</sup></u>	<u>Reduced</u> <u>R<sup>2</sup></u>
8	8	3	$X_2$	427.10		.99579	
9	9	14	$X_1 X_3^2$	470.79		.997854	
10	10	16	$X_1 X_2 X_3$	174.96		.998420	
11	11	1	$X_3$	212.37		.998899	
12	12	9	$X_2^3$	156.79		.999167	
13	13	36	$X_1^3$	230.76		.999435	
14	14	13	$X_1 X_3$	279.64		.999642	
15	15	45	$X_1^3 X_2^3$	82.42		.999694	
16	16	35	$X_1^2 X_2^3 X_3^2$	52.71		.999724	
17	17	37	$X_1^3 X_3$	107.84		.999774	
18	18	21	$X_1 X_2^3$	47.97		.999795	
19	19	39	$X_1^3 X_2$	259.26		.999867	
20	20	31	$X_1^2 X_2^2 X_3$	202.22		.999906	
25	25	23	$X_1 X_2^3 X_3^2$	76.83		.999950	
30	28	29	$X_1^2 X_2 X_3^2$	3.73		.999957	
40	30	45	$X_1^3 X_2^3$	6.72		.999960	
50	32	29	$X_1^2 X_2 X_3^2$	4.50		.999963	
60	36	41	$X_1^3 X_2 X_3^2$	10.53		.999967	
65	37	32	$X_1^2 X_2^2 X_3^2$	0.86		.999968	

#### Run 11

In this run the dependent variable, the polynomial model and the data points were the same as in Run 9. The throw-out criterion was  $F_0 = 8.0$ . The variables were included in the same order as in Run 9 through Sweep 15.

<u>Sweep</u>	<u>Terms in Model</u>	<u>Term No</u>	<u>Term</u>	<u>F in</u>	<u>F out</u>	<u>R<sup>2</sup> = Reduced R<sup>2</sup></u>
15	13	17	$x_1 x_2 x_3^2$	171.96		.999528
16	12	7	$x_2^2 x_3$		2.77	.999525
17	13	39	$x_1^3 x_2$	49.15		.999569
18	14	1	$x_3$	37.69		.999600
19	13	19	$x_1 x_2^2 x_3$		2.18	.999598
20	14	21	$x_1 x_2^3$	59.03		.999642
21	15	30	$x_1^2 x_2^2$	142.48		.999723
22	16	13	$x_1 x_3$	52.63		.999750
23	17	4	$x_2 x_3$	58.95		.999778
24	18	16	$x_1 x_2 x_3$	60.90		.999803
25	17	17	$x_1 x_2 x_3^2$		0.20	.999802
26	18	16	$x_2^3 x_3$	61.82		.999825
27	19	22	$x_1 x_2^3 x_3$	177.78		.999872
28	20	37	$x_1^3 x_3$	102.41		.999895
29	21	40	$x_1^3 x_2 x_3$	390.39		.999942
30	22	46	$x_1^3 x_2^3 x_3$	38.26		.999946
35	25	38	$x_1^3 x_3$	10.63		.999957
36	24	14	$x_1 x_3^2$		2.26	.999957
37	25	47	$x_1^3 x_2^3 x_3^2$	7.35		.999958

#### Run 12

In this run the dependent variable, the polynomial model, and the data points were the same as in Run 9. The throw-out criterion was  $F_o = 10^{-3}$ . The variables came in the same order as in Run 10 through Sweep 28.

No throw-outs were made.

<u>Sweep</u>	<u>m</u> <u>Terms</u> <u>in Model</u>	<u>Term No</u>	<u>Term</u>	<u>F in</u>	<u>F out</u>	<u>R<sup>2</sup> = Reduced R<sup>2</sup></u>
25	25	23	$X_1 X_2^3 X_3^2$	76.83		.999950
30	30	42	$X_1^3 X_2^2$	6.11		.999958
35	35	43	$X_1^3 X_2^2 X_3$	15.94		.999962
40	40	28	$X_1^2 X_2^2 X_3$	21.80		.999967
45	45	25	$X_1^2 X_3$	28.42		.999971
46	46	34	$X_1^2 X_2^3 X_3$	11.16		.999972
47	47	47	$X_1^3 X_2^3 X_3^2$	1.10		.999972

### Run 13

In this run the dependent variable was  $F(X_1, X_2, X_3) = \exp(-X_1^2 X_2 X_3)$  as in Run 9. The polynomial model was the same 47-term balanced polynomial cubic in  $X_1$  and  $X_2$ , quadratic in  $X_3$ . There were 1000 data points in a rectangular design  $X_1 = .25(.25) 2.50$ ,  $X_2 = .25(.25) 2.50$ ,  $X_3 = .25(.25) 2.50$ .

On this run the throw-out criterion was  $F_0 = 1.0$ .

<u>Sweep</u>	<u>m</u> <u>Terms</u> <u>in Model</u>	<u>Term No</u>	<u>Term</u>	<u>F in</u>	<u>F out</u>	<u>R<sup>2</sup></u>
1	1	12	$X_1$	1277.16		.561
2	2	4	$X_2 X_3$	619.57		.729
3	3	24	$X_1^2$	582.22		.829
4	4	8	$X_2^2 X_3^2$	472.27		.884
5	5	28	$X_1^2 X_2 X_3$	267.53		.9088
6	6	16	$X_1 X_2 X_3$	69.37		.9147
7	7	40	$X_1^3 X_2 X_3$	225.79		.9305

<u>Sweep</u>	<u>m</u> <u>Terms</u> <u>in Model</u>	<u>Term No</u>	<u>Term</u>	<u>F in</u>	<u>F out</u>	<u>R<sup>2</sup></u>
8	8	36	$x_1^3$	27.10		.9324
9	9	11	$x_2^3 x_3^2$	27.63		.9342
10	10	10	$x_2^3 x_3$	151.95		.9430
11	11	9	$x_2^3$	234.89		.9539
12	12	13	$x_1 x_3$	23.34		.9550
13	13	15	$x_1 x_2$	141.45		.9606
14	14	44	$x_1^3 x_2^2 x_3^2$	84.24		.9637
15	15	20	$x_1 x_2^2 x_3^2$	257.55		.9713
16	16	14	$x_1 x_3^2$	130.45		.9746
17	15	12	$x_1$		0.01	.9746
18	16	18	$x_1 x_2^2$	307.20		.980673
19	17	21	$x_1 x_2^3$	113.99		.982683
20	16	4	$x_2 x_3$		0.53	.982674
21	17	32	$x_1^2 x_2^2 x_3^2$	20.32		.983025
22	16	44	$x_1^3 x_2^2 x_3^2$		0.09	.983023
23	17	12	$x_1$	31.96		.983658
24	18	4	$x_2 x_3$	53.90		.984415
25	17	9	$x_2^3$		0.75	.984403
30	22	19	$x_1 x_2^2 x_3$	8.60		.986977
35	25	44	$x_1^3 x_2^2 x_3^2$	6.49		.988926
36	26	22	$x_1 x_2^3 x_3$	3.53		.988966
37	27	1	$x_3$	2.05		.988989

# Run 14

In this run the dependent variable, the polynomial model and the data points were the same as in Run 13. The throw-out criterion was  $F_0 = 10^{-3}$ . The tabulation is identical with Run 13 through Sweep 16.

<u>Sweep</u>	<u>Terms</u> <u>in Model</u>	<u>Term No</u>	<u>Term</u>	<u>F in</u>	<u>F out</u>	<u>R<sup>2</sup></u>
16	16	14	$X_1 X_3^2$	130.45		.9746
17	17	18	$X_1 X_2^2$	318.07		.98084
18	18	32	$X_1^2 X_2^2 X_3^2$	144.76		.98330
19	19	21	$X_1 X_2^3$	155.92		.98560
20	20	47	$X_1^3 X_2^3 X_3^2$	16.89		.98584
21	21	34	$X_1^2 X_2^3 X_3$	53.87		.98658
22	22	19	$X_1 X_2^2 X_3$	4.65		.98664
23	23	17	$X_1 X_2 X_3^2$	25.46		.98698
24	24	33	$X_1^2 X_2^3$	89.60		.98808
25	25	23	$X_1 X_2^3 X_3^2$	15.91		.98827
30	30	2	$X_3^2$	10.28		.98926
35	35	38	$X_1^3 X_3^2$	4.23		.98950
40	36	22	$X_1 X_2^3 X_3$	0.91		.989603
45	37	40	$X_1^3 X_2^3 X_3$		0.00	.989686
50	40	26	$X_1^2 X_3^2$	2.43		.989835
55	39	21	$X_1 X_2^3$		0.00	.989862
60	42	?		4.51		.990066
65	45	?		0.05		.990040
66	46	?		0.00		.990040

### Run 15

In this run, the data were taken from Bulletin 336, Agricultural Experiment Station, Auburn University, Auburn, Alabama.

The throw out was  $F_0 = 10^{-3}$  but was never used.

<u>Sweep</u>	<u>m</u> <u>Terms</u> <u>in Model</u>	<u>Term No</u>	<u>Term</u>	<u>F in</u>	<u>R<sup>2</sup></u>	<u>Reduced</u> <u>R<sup>2</sup></u>
1	1	4	$X_4$	86.98	.696	.696
2	2	2	$X_2$	3.14	.720	.712
3	3	5	$X_4^2$	0.64	.725	.710
4	4	3	$X_3$	0.24	.726	.704
5	5	6	$X_1 X_4$	0.13	.728	.696
6	6	1	$X_1$	0.58	.732	.693

### Run 16

This run used the same data as in Run 15, but the polynomial model was taken to be a balanced polynomial linear in  $X_1$ ,  $X_2$ , and  $X_3$  and quadratic in  $X_4$ . This gives 23 terms in addition to the constant term.

<u>Sweep</u>	<u>m</u> <u>Terms</u> <u>in Model</u>	<u>Term No</u>	<u>Term</u>	<u>F in</u>	<u>F out</u>	<u>R<sup>2</sup></u>	<u>Reduced</u> <u>R<sup>2</sup></u>
1	1	1	$X_4$	86.98		.69596	.69596
2	2	5	$X_3 X_4^2$	3.38		.72142	.71409
3	3	3	$X_3$	0.41		.72453	.70964
4	4	7	$X_2 X_4$	1.13		.73314	.71091
5	5	14	$X_1 X_4^2$	1.71		.74590	.71686
6	6	12	$X_1$	5.75		.78361	.75179



<u>Sweep</u>	<u>m</u> <u>Terms</u> <u>in Model</u>	<u>Term No</u>	<u>Term</u>	<u>F in</u>	<u>F out</u>	<u>R<sup>2</sup></u>	<u>Reduced</u> <u>R<sup>2</sup></u>
7	7	9	$X_2 X_3$	0.55		.78724	.74856
8	8	16	$X_1 X_3 X_4$	2.55		.80340	.76040
9	9	2	$X_4^2$	1.51		.81280	.76449
10	10	23	$X_1 X_2 X_3 X_4^2$	3.49		.83293	.78281
11	11	4	$X_3 X_4$	0.34		.83491	.77798
12	12	6	$X_2$	0.98		.840665	.778069
13	13	17	$X_1 X_3 X_4^2$	0.74		.845059	.776197
14	14	13	$X_1 X_4$	2.02		.856626	.784939
15	13	23	$X_1 X_2 X_3 X_4^2$		0.00	.856624	.792901
16	14	8	$X_2 X_4^2$	0.64		.860176	.790265
17	15	15	$X_1 X_3$	0.27		.861736	.784308
18	16	21	$X_1 X_2 X_3$	0.10		.862309	.776253
19	17	10	$X_2 X_3 X_4$	0.35		.864448	.770151
20	16	12	$X_1$		0.00	.864446	.779725
21	17	18	$X_1 X_2$	0.50		.867472	.775279
22	18	23	$X_1 X_2 X_3 X_4^2$	0.89		.872836	.774572
23	17	21	$X_1 X_2 X_3$		0.00	.872835	.784373
24	18	19	$X_1 X_2 X_4$	0.61		.876437	.780957
25	17	3	$X_3$		0.00	.876435	.790476
26	18	11	$X_2 X_3 X_4^2$	0.61		.879902	.787099
27	19	12	$X_1$	0.12		.880637	.778325
28	20	22	$X_1 X_2 X_3 X_4$	0.19		.881802	.769515

<u>Sweep</u>	<u>m</u> <u>Terms</u> <u>in Model</u>	<u>Term No</u>	<u>Term</u>	<u>F in</u>	<u>F out</u>	<u>R<sup>2</sup></u>	<u>Reduced</u> <u>R<sup>2</sup></u>
29	21	3	X <sub>3</sub>	0.20		.883701	.761282
30	20	19	X <sub>1</sub> X <sub>2</sub> X <sub>4</sub>		0.00	.883697	.773210
31	21	20	X <sub>1</sub> X <sub>2</sub> X <sub>4</sub> <sup>2</sup>	0.13		.884550	.760023
32	22	21	X <sub>1</sub> X <sub>2</sub> X <sub>3</sub>	0.06		.884958	.750743
33	23	19	X <sub>1</sub> X <sub>2</sub> X <sub>4</sub>	0.01		.885035	.736256
34	22	18	X <sub>1</sub> X <sub>2</sub>		0.00	.885034	.750906
35	23	18	X <sub>1</sub> X <sub>2</sub>	0.00		.885035	.736256

#### Run 17

In this run the data were a correlation matrix taken from Anderson, H. E., and Fruchter, B., "Predictor Selection Methods," Psychometrika, Vol. 25, No. 1, March 1960. In this run the throw-out criterion of  $F_0 = 10^{-3}$  was never used.

<u>Sweep</u>	<u>m</u> <u>Terms</u> <u>in Model</u>	<u>Term No</u>	<u>F in</u>	<u>R<sup>2</sup></u>	<u>Reduced</u> <u>R<sup>2</sup></u>
1	1	6	56.94	.156025	.156025
2	2	4	21.38	.210965	.208403
3	3	3	10.18	.236372	.231397
4	4	13	4.90	.248451	.241083
5	5	12	4.13	.258529	.248805
6	6	10	1.38	.261881	.249741
7	7	1	0.92	.264125	.249553
8	8	8	0.71	.265861	.248844

<u>Sweep</u>	<u>m Terms in Model</u>	<u>Term No</u>	<u>F in</u>	<u>R<sup>2</sup></u>	<u>Reduced R<sup>2</sup></u>
9	9	2	0.42	.266898	.247413
10	10	5	0.37	.267803	.245837
11	11	9	0.40	.268785	.244330
12	12	7	0.29	.269503	.242538
13	13	11	0.17	.269932	.240435
14	14	14	0.02	.269970	.237908

### Conclusions

We feel that the step-up procedure is an effective tool in the problem of finding a regression equation with a small number of estimation variables from a model with a large number. Using the various throw-out criteria and stopping rules, the problems of interest could be explored. The throw-out criterion and stopping rule which best fit the problem could be selected and then a regression equation determined. We feel that most future investigation of this procedure should be problem-oriented. We need the data for a problem to help develop an effective way of handling the data.

SELECTION OF SIGNIFICANT ESTIMATION VARIABLES  
IN A LEAST SQUARES PROBLEM: COMPUTER PROGRAMS

1. Comparison of variables selected by step-up procedure with optimal set. This procedure was programmed in the ALGOL 58 compiler language for the Burroughs 220 computer. Because of limitations on the memory the procedure is restricted to 25 variables.

The purpose of the program is to determine whether or not the step-up procedure actually selects the best  $k$  estimation variables. This program was preliminary to a more elaborate program for the Burroughs 5000.

First, the data are generated. The estimation variables  $Z_1, \dots, Z_{n-1}$  are terms of a balanced polynomial in independent variables  $X_1, \dots, X_\pi$ , i.e.,

$$Z_k = X_1^{l_1} \dots X_\pi^{l_\pi}, \quad l_i = 0, \dots, L_i, \quad i=1, 2, \dots, \pi,$$

where  $(l_1, \dots, l_\pi)$  takes on all possible values in the given range except  $(0, \dots, 0)$ . Certain terms of the balanced polynomial are to be used to estimate a dependent variable, which is some function of the  $X$ 's. It is convenient to label this variable  $Z_n$ . Corresponding to an index,  $t_i = 1, 2, \dots, T_i, i=1, 2, \dots, \pi$ , the observed value of  $X_i$  is  $x_{it_i}$ . Thus, corresponding to the set  $\{(t_1, \dots, t_\pi) | t_i = 1, 2, \dots, T_i, i=1, 2, \dots, \pi\}$  is a rectangular set of data-points  $\{(x_{1t_1}, \dots, x_{\pi t_\pi})\}$  from which are calculated observed values,  $(z_{\mu 1}, \dots, z_{\mu, n-1}, z_{\mu n})$ , of the vector consisting of the estimation variables and the dependent variable.

Next, regression analyses are made using all possible combinations of  $k$  estimation variables, where  $k=2, \dots, n-2$ . For each  $k$ , the combinations of variables which give maximum and minimum sums of squares due to regression (and hence maximum and minimum multiple correlation) are printed along with the sums of squares.

Finally, the step-up procedure is used. At the  $k$ 'th step, the variable is selected from those not already included which maximizes  $S_{kn}^{(k')^2} / S_{kk}^{(k')}$ . The procedure then uses that variable  $Z_k$ , as the pivot variable. It makes the following calculations:

$$S_{kj}^{(k'+1)} = \frac{S_{kj}^{(k')}}{S_{kk}^{(k')}} \quad j = 1, 2, \dots, n$$

$$S_{ij}^{(k'+1)} = S_{ij}^{(k')} - \frac{S_{ik}^{(k')} S_{kj}^{(k')}}{S_{kk}^{(k')}} \quad i = 1, \dots, k-1, k+1, \dots, n, j = 1, 2, \dots, n.$$

In these calculations  $(S_{ij})$  is the augmented matrix of dot products of the estimation vectors and the dependent-variable vector. The superscript  $k'$  indicates the number of transformations on  $(S_{ij})$  in which a column has been reduced to a unit vector. The list of variables, included in the regression, and the sum of squares due to regression are printed.

In some cases the stepwise procedure gave optimal solutions, while in others it did not. In an attempt to run the program with 18 variables the time required to calculate the regression analyses for all combinations of variables turned out to be prohibitive.

#### Operating Instructions for B-220 Program

1. Load the program, with the proper procedure (FCN) inserted to calculate the independent polynomial variables and the dependent variables.
2. Load the following data card, using more than one card if necessary, with 5 punched in the first column of each card.

Card Contents	Card Format
a) Number of independent polynomial variables	Skip at least one column; punch integer
b) Number of observations of independent polynomial variable	Skip at least one column; punch integer
c) Repeat (b) for each variable	
d) Order in independent polynomial variable	Skip at least one column; punch integer
e) Repeat (d) for each variable	
f) Lower bound for diagonal element	Skip at least one column; punch floating point number
g) Lower bound for difference between 1.0 and off-diagonal correlation	Skip at least one column; punch floating point number
h) F-statistic for stopping	Skip at least one column; punch floating point number; leave rest of card blank.

3. Repeat (2) for each analysis to be made.

4. Load 2 blank cards.

2. Comprehensive program for selection of variables with step-up procedure incorporating elimination rules and stopping rules. This procedure attempts to select the most significant estimation variables for a least squares fitting. It has been programmed for the Burroughs 5000 computer in the ALGOL 60 compiler language.

There are ~~three~~ options for obtaining the  $n \times n$  augmented  $(S_{ij})$  matrix

(1) Either the  $(S_{ij})$  matrix or the correlation matrix may be read in. (Only the diagonal and lower triangle are read in.)

(2) Each of the  $M$  observations  $(z_{\mu 1}, \dots, z_{\mu n})$  may be read in. An estimate  $(m_1, \dots, m_n)$  of the means is available. As the data are read

in, the sums

$$S_i = \sum_{\mu=1}^M (Z_{\mu i}^{-m_i})$$

$$S'_{ij} = \sum_{\mu=1}^M (Z_{\mu i}^{-m_i}) (Z_{\mu j}^{-m_j}) \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, i$$

are calculated. The adjusted  $(S_{ij})$  matrix

$$S_{ij} = S'_{ij} - \frac{S_i S_j}{M} \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, i$$

is then computed.

(3) Each observation may be generated from balanced polynomials. A set of fixed data points  $(x_{\mu 1}, \dots, x_{\mu \pi})$  is given. The estimation variables are the terms of a balanced polynomial, so that

$$z_{\mu k} = x_{\mu 1}^{l_1} x_{\mu 2}^{l_2} \dots x_{\mu \pi}^{l_\pi}$$

where  $l_i = 0, 1, \dots, L_i$ ,  $i = 1, 2, \dots, \pi$ . Each of these combinations of exponents (except all exponents zero) corresponds to one estimation variable. The values  $x_{\mu 1}, \dots, x_{\mu \pi}$  may be read in, or they may be part of a rectangular design, with each  $\mu$  corresponding to some value of the index  $(t_1, \dots, t_\pi)$ , where  $t_i = 1, \dots, T_i$ ,  $i = 1, 2, \dots, \pi$ . Values  $z_{\mu n}$  of the dependent variable may be read in or they may be computed

values of a specified function, corresponding to values  $x_{\mu 1}, \dots, x_{\mu n}$ . These vectors  $x_{\mu 1}, \dots, x_{\mu n}$  are generated in a procedure which may be varied with each run. As the observations  $z_{\mu 1}, \dots, z_{\mu n}$  are generated, the sum of squares matrix  $(S_{ij})$  is calculated as above.

Once the adjusted sum of squares matrix has been obtained it may be used for more than one analysis. The diagonal and lower triangle only are used in the analysis. Since the matrix is symmetric, the necessary values may be stored in the upper triangle (with the diagonal in a separate vector) for performing other analyses under different conditions.

If the correlation matrix was read in, it is used in the regression analysis; otherwise, there is the option of computing and using the correlation matrix. The matrix to be used shall be denoted as  $(S_{ij}^{(0)})$ . The program includes the option of printing this matrix.

In a hand computation the system of normal equations would be solved for regression coefficients in a sequence of gaussian eliminations, and the inverse matrix would be built up on a unit matrix. The initial tableau  $(R_{ij}^{(0)})$  for such an elimination and matrix inversion procedure would be defined by



$$\hat{R}_{ij}^{(0)} = \begin{cases} s_{ij}^{(0)} & i = 1, 2, \dots, n; j = 1, 2, \dots, i \\ s_{ji}^{(0)} & i = 1, 2, \dots, n-1; j = i+1, \dots, n \\ 1 & i = 1, 2, \dots, n; j = n+1 \\ 0 & i = 1, 2, \dots, n; j = n+1, \dots, 2n, j \neq n+i \end{cases}$$

The original S matrix is of the form

$$\begin{pmatrix} s_{11}^{(0)} & & & \\ s_{21}^{(0)} & s_{22}^{(0)} & & \\ \vdots & \vdots & \ddots & \\ s_{n-1,1}^{(0)} & s_{n-1,2}^{(0)} & \dots & s_{n-1,n-1}^{(0)} \\ s_{n1}^{(0)} & s_{n2}^{(0)} & \dots & s_{n,n-1}^{(0)} & s_{nn}^{(0)} \end{pmatrix}$$

while the original R matrix is of the form

$$\begin{array}{cccccccc}
s_{11}^{(0)} & s_{21}^{(0)} & \dots & s_{n-1,1}^{(0)} & s_{n1}^{(0)} & 1 & 0 & \dots & 0 \\
s_{21}^{(0)} & s_{22}^{(0)} & \dots & s_{n-1,2}^{(0)} & s_{n2}^{(0)} & 0 & 1 & \dots & 0 \\
\vdots & \vdots & & & \vdots & \vdots & & & \vdots \\
s_{n-1,1}^{(0)} & s_{n-1,2}^{(0)} & \dots & s_{n-1,n-1}^{(0)} & s_{n,n-1}^{(0)} & 0 & 0 & \dots & 1 & 0 \\
s_{n1}^{(0)} & s_{n2}^{(0)} & \dots & s_{n,n-1}^{(0)} & s_{nn}^{(0)} & 0 & 0 & \dots & 0 & 1
\end{array}$$

Because of symmetry operations need to be made only on the lower triangle of the S matrix. Hence the entire R matrix need not be stored in memory.

The stepwise procedure now begins. It is assumed that at the k'th step, k estimation variables  $Z_{p_1}, \dots, Z_{p_k}$  are included in the regression, while the n-k-l variables  $Z_{q_1}, \dots, Z_{q_{n-k-l}}$  are excluded. The variables

$Z_{p_{\max}}$  and  $Z_{q_{\min}}$  which minimize  $(s_{np_i}^{(k')^2})/s_{p_i p_i}^{(k')}$  and maximize  $(s_{nq_j}^{(k')^2})/s_{q_j q_j}^{(k')}$ , respectively, are determined. The variable  $Z_{p_{\min}}$  shall be considered significant if

$$\frac{(S_{np_{\min}}^{(k')})^2 / S_{p_{\min} p_{\min}}^{(k')}}{S_{nn}^{(k')} / (M-k-1)} \geq F_0$$

and the variable  $Z_{q_{\max}}$  shall be considered significant if

$$\frac{(S_{nq_{\max}}^{(k')})^2 / S_{q_{\max} q_{\max}}^{(k')}}{[S_{nn}^{(k')} - (S_{nq_{\max}}^{(k')})^2 / S_{q_{\max} q_{\max}}^{(k')}] / (M-k-2)} > F_I$$

where  $F_I$  and  $F_0$  are criteria based on the F-distribution.  $F_I$  should not be less than  $F_0$ ; if it were, looping might occur.

The procedure now tests whether  $Z_{p_{\min}}$  is to be dropped from the regression. There are two options for dropping a variable:

- (1) If  $Z_{p_{\min}}$  is not significant, it is dropped. (This may be bypassed by setting  $F_0$  equal to zero.)
- (2) The procedure alternately adds two variables and drops one.  
If  $Z_{p_{\min}}$  is not to be dropped, the procedure checks whether to stop or not.

There are four criteria for stopping, the first two of which are now checked.

- (1) If  $Z_{p_{\max}}$  is not significant, it is added and then the procedure terminates. (This may be bypassed by setting  $F_I$  to zero.)
- (2) When a specified maximum number of terms have been included in

the regression, the procedure terminates. Unless otherwise specified, this will be the number of estimation variables.

- (3) If the square of the multiple correlation coefficient is greater than a specified amount  $R_{\max}^2$ , the procedure terminates. (This may be bypassed by setting  $R_{\max}^2$  to 1.)
- (4) When the procedure has gone through a specified number of iterations, it terminates. If the procedure is following the option of adding two variables and dropping one, this will be three times the maximum number of terms; otherwise, it will be twice the maximum number of terms.

If  $Z_{p_{\min}}$  is not to be dropped, and if the procedure does not stop,  $Z_{q_{\max}}$  is now added to the regression.

The  $j$ th column of the  $S$  matrix corresponds to the  $(j+n)$ -th of the  $R$  matrix if the  $j$ th variable has been included in the regression and to the  $j$ th column otherwise. (At all stages, either the  $j$ th column or the  $(j+n)$ -th column of the  $R$  matrix will be a unit vector. The  $S$  matrix will contain the column which is not. Of course the storage of the unit vector is unnecessary.)

It will be assumed that the  $q$ th variable is to be added or dropped. (The computational procedure is the same in both cases. It will also be assumed that  $H_j^{(k')} = -1$  if the  $j$ -th variable is included in the regression after  $k'$  iterations and that  $H_j^{(k')} = +1$  otherwise. Note that  $H_n^{(k')} = +1$  throughout the analysis.  $H_q^{(k')}$  depends on the status of the  $q$ th variable before, rather than after it is added or dropped.)

The following formulae determine the matrix  $(S_{ij}^{(k'+1)})$  :

$$S_{qq}^{(k'+1)} = \frac{1}{S_{qq}^{(k')}}}$$

$$S_{qj}^{(k'+1)} = \frac{S_{qj}^{(k')}}{S_{qq}^{(k')}}} \quad j < q$$

$$S_{iq}^{(k'+1)} = - \frac{S_{iq}^{(k')}}{S_{qq}^{(k')}}} \quad i > q$$

$$S_{ij}^{(k'+1)} = S_{ij}^{(k')} - \frac{S_{qi}^{(k')} S_{qj}^{(k')} H_i^{(k')} H_q^{(k')}}{S_{qq}^{(k')}}} \quad j \leq i < q$$

$$S_{ij}^{(k'+1)} = S_{ij}^{(k')} - \frac{S_{iq}^{(k')} S_{qj}^{(k')}}{S_{qq}^{(k')}}} \quad j < q < i$$

$$S_{ij}^{(k'+1)} = S_{ij}^{(k')} - \frac{S_{iq}^{(k')} S_{jq}^{(k')} H_j^{(k')} H_q^{(k')}}{S_{qq}^{(k')}}} \quad q < j \leq i$$

This is equivalent, on adding a variable, to

$$R_{qj}^{(k'+1)} = \frac{R_{qj}^{(k')}}{R_{qq}^{(k')}}}$$

$$R_{ij}^{(k'+1)} = R_{ij}^{(k'+1)} - \frac{R_{iq}^{(k')} R_{qj}^{(k')}}{R_{qq}^{(k')}}}$$

or, on dropping a variable, to

$$R_{qj}^{(k'+1)} = \frac{R_{qj}^{(k')}}{R_{q,q+n}^{(k')}}^{(k')}$$

$$R_{ij}^{(k'+1)} = R_{ij}^{(k')} - \frac{R_{i,q+n}^{(k')} R_{qj}^{(k')}}{R_{q,q+n}^{(k')}}^{(k')}$$

where the  $(q+n)$ -th column of the R matrix takes the place of the  $q$ th in the S matrix when a variable is being added.

If the first  $k$  variables were included in the regression, then the R matrix would be of the form.

$$\begin{bmatrix} 1 & 0 & -S_{k+1,l}^{(k)} & \dots & -S_{n-l,l}^{(k)} & -S_{nl}^{(k)} & S_{1l}^{(k)} & \dots S_{kl}^{(k)} & 0 & \dots & 0 \\ \dots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \dots \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & -S_{k+1,k}^{(k)} & \dots & -S_{n-l,k}^{(k)} & -S_{nk}^{(k)} & S_{kl}^{(k)} & \dots S_{kk}^{(k)} & 0 & \dots & 0 \\ 0 & \dots & 0 & S_{k+1,k+1}^{(k)} & \dots & S_{n-l,k+1}^{(k)} & S_{n,k+1}^{(k)} & S_{k+1,l}^{(k)} & \dots S_{k+1,k}^{(k)} & 1 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \vdots & \dots & \dots & \vdots \\ 0 & \dots & 0 & S_{n-l,k+1}^{(k)} & S_{n-l,n-l}^{(k)} & S_{n,n-l}^{(k)} & S_{n-l,l}^{(k)} & \dots S_{n-l,k}^{(k)} & 0 & 1 & 0 \\ 0 & \dots & 0 & S_{n,k+1}^{(k)} & S_{n,n-l}^{(k)} & S_{nn}^{(k)} & S_{nl}^{(k)} & \dots S_{nk}^{(k)} & 0 & \dots & 0 \end{bmatrix}$$

In effect the program inverts the S matrix in place, proceeding from pivot element to pivot element without rearranging rows and columns. Also,

advantage is taken of the symmetry in carrying out calculations in the lower triangle only.

At this point, a list of included or active variables, the mean-squares due to regression and to error, the F-ratio, and the square of the multiple correlation coefficient are printed. There are options for printing the inverse matrix, the reduced sum of squares matrix, the partial regression coefficients of the dependent variable on each of the active variables, and the regression coefficients of the dependent variable on the active variables.

## REFERENCES

### On Least Squares:

GRAYBILL, F.A., An Introduction to Linear Statistical Models, McGraw-Hill Co., Inc., N.Y., 1961.

SCARBOROUGH, J.B., Numerical Mathematical Analysis, John Hopkins Press, Baltimore, 1958.

SCHEFFE, Henri, The Analysis of Variance, John Wiley and Sons, Inc., N.Y., 1959.

WILLIAMS, E.J., Regression Analysis, John Wiley and Sons, Inc., N.Y., 1959.

ZELEN, Marvin, "Linear Estimation and Related Topics," A Survey of Numerical Analysis, edited by John Todd, McGraw-Hill and Co., Inc., N.Y., 1962.

### On the Step-up Procedure:

ANDERSON, Harry E., and B. FRUCHTER, "Some Multiple Correlation and Predictor Selection Methods," Psychometrika, 25 (Mar. 1960), 59-76.

SCHULTZ, E.F., Jr., and J.F. GOGGANS, "A Systematic Procedure in Determining Potent Independent Variables in Multiple Regression and Discussion of Analysis" Bulletin 336, Agricultural Experiment Station, Auburn University, November, 1961.

WHERRY, R.J., "A New Formula for Predicting the Shrinkage of the Coefficient of Multiple Correlation," Annals of Math. Stat., 2, (1931), 440-451.



SOUTHERN ILLINOIS UNIVERSITY  
Carbondale, Illinois

ON THE NUMERICAL REPRESENTATION OF THE GENERAL  
SOLUTION OF SYSTEMS OF ORDINARY DIFFERENTIAL EQUATIONS

By

Robert Silber

M65 33064

# ON THE NUMERICAL REPRESENTATION OF THE GENERAL SOLUTION OF SYSTEMS OF ORDINARY DIFFERENTIAL EQUATIONS

By

Robert Silber

## I. INTRODUCTION AND SUMMARY

We consider normal systems of first order, ordinary, differential equations, i.e., we consider the system

$$\dot{y}_i = f_i(t, y_1, y_2, \dots, y_n), \quad i = 1, 2, \dots, n, \quad (s)$$

in which the dot indicates differentiation with respect to  $t$ .  
Let the set of functions

$$Y_i(t, \tau, \eta_1, \eta_2, \dots, \eta_n), \quad i = 1, 2, \dots, n,$$

be the general solution to (s) in terms of the initial time  $\tau$  and the initial values  $\eta_i$  of the  $y_i$ .

Under certain conditions, such as those discussed, the functions  $Y_i$  will be analytic at a selected point  $(t^*, \tau^*, \eta_1^*, \eta_2^*, \dots, \eta_n^*)$  and will therefore be expressible in Taylor's series in  $n+2$  variables neighboring the point  $(t^*, \tau^*, \eta_1^*, \eta_2^*, \dots, \eta_n^*)$ . The information needed to calculate the coefficients in this Taylor's series is the set of values of the partial derivatives of the  $Y_i$  at the point  $(t^*, \tau^*, \eta_1^*, \eta_2^*, \dots, \eta_n^*)$ .

Within the numerical procedures discussed in Reference 3, there is contained a method for obtaining the values of the above partial derivatives, through any pre-specified order. The method necessitates the use of a digital computer. In writing Reference 3, this method was not given explicit mention, because it was integrated into a more complex numerical process. Since writing Reference 3, the author has come to realize that perhaps the subject method is of sufficient interest to merit an independent description. Thus, the purpose and content of this paper is a description of the salient points of this method; many of the troublesome details and minutiae are left untreated, since they are all contained in Reference 3.

## II. THE FUNDAMENTAL IDENTITIES

The entire procedure is based on two fundamental identities satisfied by the functions  $Y_i$ . Before writing the identities, it will be convenient to introduce abbreviated notation as follows:

$$Y = (Y_1, Y_2, \dots, Y_n),$$

$$\eta = (\eta_1, \eta_2, \dots, \eta_n).$$

Thus  $Y(t, \tau, \eta) = (Y_1(t, \tau, \eta_1, \eta_2, \dots, \eta_n), \dots, Y_n(t, \tau, \eta_1, \eta_2, \eta_n))$ .

The first of the fundamental identities is a consequence of the  $Y_i$  being solutions to (s).

$$\boxed{\begin{aligned} \frac{\partial}{\partial t} Y_i(t, \tau, \eta) &= f_i(t, Y(t, \tau, \eta)), \\ i &= 1, 2, \dots, n. \end{aligned}} \quad (1)$$

This is an identity in each of the  $n+2$  arguments which appear. In the event that each  $f_i$  is analytic at the point  $(t, Y(t, \tau, \eta))$  and each  $Y_i$  is analytic at the point  $(t, \tau, \eta)$ , the two sides of (1) represent the same analytic function, and new identities can be obtained from (1) by unlimited differentiation. Thus, for example, using the chain rule,

$$\begin{aligned} \frac{\partial^2 Y_i}{\partial t^2}(t, \tau, \eta) &= \frac{\partial f_i}{\partial t}(t, Y(t, \tau, \eta)) \\ &+ \sum_{j=1}^n \frac{\partial f_i}{\partial y_j}(t, Y(t, \tau, \eta)) \frac{\partial Y_j}{\partial t}(t, \tau, \eta), \end{aligned}$$

which, using (1), can be written

$$\boxed{\begin{aligned} \frac{\partial^2 Y_i}{\partial t^2}(t, \tau, \eta) &= \frac{\partial f_i}{\partial t}(t, Y(t, \tau, \eta)) \\ &+ \sum_{j=1}^n f_j(t, Y(t, \tau, \eta)) \frac{\partial f_i}{\partial y_j}(t, Y(t, \tau, \eta)), \\ i &= 1, 2, \dots, n. \end{aligned}} \quad (2)$$

Similarly,

$$\frac{\partial^2 Y_i}{\partial t \partial \tau} (t, \tau, \eta) = \sum_{j=1}^n \frac{\partial f_i}{\partial y_j} (t, Y(t, \tau, \eta)) \frac{\partial Y_j}{\partial \tau} (t, \tau, \eta), \quad (3)$$

$$i = 1, 2, \dots, n,$$

and

$$\frac{\partial^2 Y_i}{\partial t \partial \eta_k} (t, \tau, \eta) = \sum_{j=1}^n \frac{\partial f_i}{\partial y_j} (t, Y(t, \tau, \eta)) \frac{\partial Y_j}{\partial \eta_k} (t, \tau, \eta), \quad (4)$$

$$i, k = 1, 2, \dots, n.$$

Clearly, by repeated differentiations, one can obtain identities involving partial derivatives of higher orders.

The second of the two fundamental identities is a consequence of the definition of the parameters  $\tau, \eta_1, \eta_2, \dots, \eta_n$  as being "initial values."

$$Y_i(\tau, \tau, \eta) = \eta_i, \quad i=1, 2, \dots, n. \quad (5)$$

As in (1), this is an identity in each of the  $n+1$  arguments appearing, and both sides can be differentiated indefinitely at points of analyticity. Hence,

$$\frac{\partial Y_i}{\partial t} (\tau, \tau, \eta) + \frac{\partial Y_i}{\partial \tau} (\tau, \tau, \eta) = 0,$$

so that by (1) and (5),

$$\frac{\partial Y_i}{\partial \tau} (\tau, \tau, \eta) = -f_i(\tau, Y(\tau, \tau, \eta)) = f_i(\tau, \eta) \quad (6)$$

$$i = 1, 2, \dots, n.$$

Also,

$$\frac{\partial Y_i}{\partial \eta_k}(\tau, \tau, \eta) = \delta_{ik}; \quad i, k = 1, 2, \dots, n, \quad (7)$$

where  $\delta_{ik}$  is the Kronecker delta.

Again, as in the case of Equation (1), further differentiations can be performed, yielding identities involving partial derivatives of progressively higher orders.

In the procedure to follow, Equations (1)-(7), and higher order equations to be obtained through appropriate differentiations, will be used.

### III. REFERENCE POINTS AND REFERENCE TRAJECTORIES

As was pointed out in the introduction, the aim of the method being described is the expansion of the functions  $Y_i$ ,  $i=1,2,\dots,n$ , in Taylor's series about the pre-specified point  $(t^*, \tau^*, \eta_1^*, \eta_2^*, \dots, \eta_n^*)$ . It is a clear necessity that the functions  $Y_i$  be analytic at this point. Analyticity is also sufficient for existence and convergence of the Taylor's series neighboring the point of expansion, but for our method we shall require further properties. To facilitate the discussions concerned with these properties, we introduce some definitions.

Definition: A real solution of (s) over a real interval  $[a,b]$  is a set  $\{\varphi_1, \varphi_2, \dots, \varphi_n\}$  of real-valued functions, defined and differentiable on  $[a,b]$ , and satisfying

$$\dot{\varphi}_i(t) = f_i(t, \varphi_1(t), \varphi_2(t), \dots, \varphi_n(t),$$

$$i = 1, 2, \dots, n; \quad t \in [a,b].$$

In keeping with our earlier abbreviated notation, we let  $\varphi = (\varphi_1, \varphi_2, \dots, \varphi_n)$ ,  $f = (f_1, f_2, \dots, f_n)$ , and write the above equation

$$\dot{\varphi}(t) = f(t, \varphi(t)), \quad t \in [a,b],$$

where, of course,  $\dot{\varphi} = (\dot{\varphi}_1, \dot{\varphi}_2, \dots, \dot{\varphi}_n)$ . We shall refer to  $\varphi$  itself as the solution over  $[a,b]$ .

Definition: Suppose  $\varphi$  is a solution of  $s$  over  $[a,b]$  .  
The set

$$\mathcal{O}(\varphi, a, b) = \{ \varphi(t) : t \in [a, b] \} ,$$

which is a subset of  $n$ -dimensional space, is called the orbit of  $\varphi$ , over  $[a,b]$  . The set

$$\mathcal{J}(\varphi, a, b) = \{ (t, \varphi(t)) : t \in [a, b] \} ,$$

which is a subset of  $(n+1)$  dimensional space, is called the trajectory of  $\varphi$ , over  $[a,b]$ . A reference trajectory is a trajectory  $\mathcal{J}(\varphi, a, b)$  of a solution  $\varphi$  over an interval  $[a,b]$  with the following property:

At each point  $(t, \varphi(t)) \in \mathcal{J}(\varphi, a, b)$ , each of the functions  $f_i$ ,  $i=1, 2, \dots, n$ , in  $(s)$ , is analytic\*.

A real reference trajectory is a reference trajectory  $\mathcal{J}(\varphi, a, b)$  for which  $\varphi$  is real-valued in each component. Analyticity, however, is still taken in the complex sense. (cf. the definition below.)

From the theory of differential equations (References 1 and 2), it is known that if  $(\tau, \eta_1, \eta_2, \dots, \eta_n)$  is a point at which each function  $f_i$ ,  $i=1, 2, \dots, n$ , in  $(s)$ , is analytic, then there exists a unique complex function  $\varphi$  of the complex variable  $z$  which is analytic in a complex neighborhood  $N$  of  $\tau$ , which satisfies  $\varphi(\tau) = \eta$  and which solves  $(s)$  at each point of  $N$ .

Definition: A point  $(t^*, \tau^*, \eta_1^*, \eta_2^*, \dots, \eta_n^*)$  shall be called a reference point if the following conditions are met:

\*  $f_i$  is analytic at  $(t, \varphi_1(t), \varphi_2(t), \dots, \varphi_n(t))$ , if  $f_i$  is expressible by a power series

$$\sum C^{(i)}_{\nu_0 \nu_1 \nu_2 \dots \nu_n} (z_0 - t)^{\nu_0} (z_1 - \varphi_1(t))^{\nu_1} \dots (z_n - \varphi_n(t))^{\nu_n},$$

which is convergent throughout an  $(n+1)$  complex dimensional neighborhood of  $(t, \varphi(t))$ , and represents  $f_i$  there.

(i) Each  $f_i$ ,  $i=1,2,\dots,n$ , is analytic at  $(t^*, \eta_1^*, \eta_2^*, \dots, \eta_n^*)^1$  and bounded on some complex neighborhood of that point.

(ii) Let  $\varphi$  be the unique solution of (s), analytic at  $\tau^*$ , and satisfying  $\varphi(t^*) = \eta^*$ . Then  $\varphi$  has an analytic continuation  $\tilde{\varphi}$  along the real axis, from  $\tau^*$  to  $t^*$ .

(iii)  $\mathcal{J}(\tilde{\varphi}, \tau^*, t^*)$  (or  $\mathcal{J}(\tilde{\varphi}, t^*, \tau^*)$ , if  $t^* < \tau^*$ ) is a reference trajectory.

From this definition, it follows (for example, from theorem 8.2 in chapter one of Reference 1) that if

$(t^*, \tau^*, \eta_1^*, \eta_2^*, \dots, \eta_n^*)$  is a reference point, then the general solution  $Y(t, \tau, \eta_1, \eta_2, \dots, \eta_n)$  mentioned in the introduction, is well-defined and analytic at each point  $(t, \tau, \eta)$  such that  $(\tau, \eta) \in \mathcal{J}(\tilde{\varphi}, \tau^*, t^*)$  and  $t \in [\tau^*, t^*]$ , and satisfies  $Y(t, \tau^*, \eta^*) = \tilde{\varphi}(t)$ , for each  $t \in [\tau^*, t^*]$ . Thus, each of the preceding differentiations of identities is justified.

In practice, the system (s) has the property that its solutions are real if  $t$  is real, and if the initial values are real; consequently, the points in  $\mathcal{J}(\tilde{\varphi}, \tau^*, t^*)$  will have real components instead of complex components. Nevertheless, a complex, rather than real, notion of analyticity must be retained, in order to justify the differentiations on which the numerical procedure is to be based.

#### IV. NUMERICAL CALCULATIONS

Let  $(t^*, \tau^*, \eta_1^*, \eta_2^*, \dots, \eta_n^*)$  be a given reference point, and let  $\tilde{\varphi}(t) = Y(t, \tau^*, \eta^*)$ ,  $t \in [\tau^*, t^*]$ , as in the preceding definitions. Let  $\mathcal{J}^* = \mathcal{J}(\tilde{\varphi}, \tau^*, t^*)$ , the reference trajectory defined by, and corresponding to, the given reference point.

By numerical integration on a digital computer, points of  $\mathcal{J}^*$  can be calculated at selected values of  $t$  in  $[\tau^*, t^*]$ , so that we assume  $\mathcal{J}^*$  to be numerically known.

We now define an  $n \times n$  matrix function of time. Let  $F$  be the matrix with elements  $f_{ij}$  defined by

$$f_{ij}(t) = \frac{\partial f_i}{\partial y_j}(t, \tilde{\varphi}(t)); \quad t \in [\tau^*, t^*];$$

$$i, j = 1, 2, \dots, n.$$

In keeping with popular terminology, we shall call  $F$  the transition matrix of the system (s) along the trajectory  $J^*$ . Notice that since (s) is given, the functions

$$\frac{\partial f_i}{\partial y_j}(t, y_1, y_2, \dots, y_n)$$

can be obtained by direct differentiation of the right hand sides of (s). Further, since  $J^*$  is numerically known, we can assume that by further calculations,  $F$  is numerically known for each  $t \in [\tau^*, t^*]$ .

Next, we define an  $n \times (n+1)$  matrix function of time. Let  $X$  be the matrix with elements  $x_{ij}$ , defined by

$$x_{ij}(t) = \frac{\partial Y_i}{\partial \eta_j}(t, \tau^*, \eta^*); \quad t \in [\tau^*, t^*], \quad i, j = 1, 2, \dots, n,$$

$$x_{ij}(t) = \frac{\partial Y_i}{\partial \tau}(t, \tau^*, \eta^*); \quad t \in [\tau^*, t^*], \quad i = 1, 2, \dots, n, \\ j = n+1.$$

If, in Equations (3) and (4), one sets  $(\tau, \eta) = (\tau^*, \eta^*)$ , then each side of the equations becomes a function of  $t$  alone, and partial derivatives with respect to  $t$  become total. Furthermore, the equations, taken over all indices are equivalent to the single matrix equation

$$\dot{X} = FX. \quad (8)$$

Furthermore, Equations (6) and (7) give the entries of  $X(\tau^*)$ ;

$$X(\tau^*) = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & -f_1(\tau_1^*, \eta^*) \\ 0 & 1 & 0 & \dots & 0 & -f_2(\tau^*, \eta^*) \\ 0 & 0 & 1 & \dots & 0 & -f_3(\tau^*, \eta^*) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & -f_n(\tau^*, \eta^*) \end{bmatrix} \quad (9)$$

$\underbrace{\hspace{10em}}_{n\text{th column}} \quad \underbrace{\hspace{10em}}_{(n+1)\text{st column}}$



Direct numerical integration of (8) from  $t = \tau^*$  to  $t = t^*$ , using the initial value given by (9), will yield, with one exception, all first partials needed for the Taylor's series. The exception is

$$\frac{\partial y_i}{\partial t}(t^*, \tau^*, \eta^*), \quad i = 1, 2, \dots, n,$$

but this can be calculated directly from the point  $(t^*, \tilde{\phi}(t^*))$  in  $\mathcal{J}^*$  and the right side of Equation (1).

The method extends readily to higher order partials. The analogue of (8) must be obtained by differentiations of (3) and (4), and the analogue of (9) by differentiations of (6) and (7). The new matrix equations will be of higher dimensions since there are many more second partials than first partials. The involved equations and determinations are treated in detail in reference three, and so are not taken up here. It is felt that a detailed description for first-order partials is sufficient to convey the basic ideas of the method.

#### REFERENCES

1. Coddington, E. A., and Levinson, N., Theory of Ordinary Differential Equations, McGraw-Hill, 1955. pp 1-37, Esp. Theorem 8.2.
2. Dieudonne', J., Foundations of Modern Analysis, Academic Press, 1960. Chapters 9 and 10.
3. NASA TM X-53059, "Space Vehicle Guidance - A Boundary Value Formulation," by Robert Silber and Robert W. Hunt, June 8, 1964.
4. NASA TM X-53100, "Space Vehicle Guidance - A Boundary Value Formulation. Part II, Boundary Conditions with Parameters," by Robert Silber, July 28, 1964.

APPROVAL PAGE

NASA TM X-53292

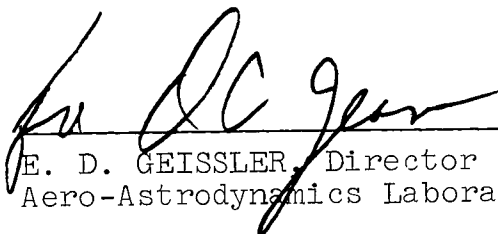
PROGRESS REPORT NO. 7

on Studies in the Fields of  
SPACE FLIGHT AND GUIDANCE THEORY

-----  
Sponsored by Aero-Astroynamics Laboratory  
of Marshall Space Flight Center

The information in this report has been reviewed for security classification. Review of any information concerning Department of Defense or Atomic Energy Commission programs has been made by the MSFC Security Classification Officer. This report, in its entirety, has been determined to be unclassified.

This document has also been reviewed and approved for technical accuracy.

  
\_\_\_\_\_  
E. D. GEISLER, Director  
Aero-Astroynamics Laboratory

Mr. T. Perkins  
Chrysler Corporation  
HIC Building  
Huntsville, Alabama

Mr. George Cherry  
Massachusetts Institute of Technology  
Cambridge, Massachusetts

Mr. C. R. Coates  
Astrodynamics Section  
Astrosciences Department  
Aeronutronic Division of Ford Motor Co.  
Ford Road  
Newport Beach, California

Mr. Walter L. Portugal  
Manager, Systems Sales  
General Precision, Inc.  
Aerospace Group  
Systems Division  
Little Falls, New Jersey

Mr. M. D. Anderson, Jr.  
General Dynamics Corporation  
Suite 42 Holiday Office Center  
South Memorial Parkway  
Huntsville, Alabama

Mr. Y. L. Luke  
Mathematics & Physics Division  
Midwest Research Institute  
425 Volker Boulevard  
Kansas City 10, Missouri

Mr. Robert Allen  
Manager, Huntsville Sales Office  
A. C. Spark Plug  
The Electronics Division  
of General Motors  
Holiday Office Center  
Huntsville, Alabama

Mr. Ted Quinn  
Advanced Space Technology  
Engineering Research  
Douglas Aircraft Corporation  
Santa Monica, California

Mr. J. W. Scheuch  
North American Aviation  
P. O. Box 557  
Huntsville, Alabama

Mr. Arthur C. Gilbert, Sc. D.  
Chief, Space Systems Requirements  
Corporate Systems Center  
Division United Aircraft Corporation  
1690 New Britain Avenue  
Farmington, Connecticut

Mr. Robert A. Lerman  
Sr. Analytical Engineer  
Technical Planning  
Hamilton Standard Division  
United Aircraft Corporation  
Windsor Locks, Connecticut

Mr. Robert G. Chilton  
NASA-Manned Spacecraft Center-EG  
P. O. Box 1537  
Houston, Texas

Mr. R. W. Reck (5)  
Martin Company  
3312 S. Memorial Parkway  
Huntsville, Alabama

Dr. Stanley E. Ross  
Chief, Advanced Program Analysis  
Advanced Manned Missions  
Systems Engineering  
National Aeronautics and  
Space Administration  
Washington, D. C.

Stephen J. Kahne  
Lt., USAF  
Applied Mathematics Branch  
Data Sciences Laboratory  
Air Force Cambridge Research Laboratories  
Office of Aerospace Research  
Lawrence G. Hanscom Field  
Bedford, Massachusetts

Dr. Bob Plunkett  
Northrop Corporation  
Sparkman Drive  
Huntsville, Alabama

Mr. J. P. deVries  
Astrodynamics Operation  
Space Sciences Laboratory  
Missile and Space Vehicle Department  
General Electric Company  
Valley Forge Space Technology Center  
P. O. Box 8555  
Philadelphia 1, Pennsylvania

Mr. William R. Wells  
Mail Stop #304  
Langley Research Center  
Space Mechanics Division  
Langley Field Air Force Base  
Hampton, Virginia

Mr. R. M. Chapman  
Lockheed Missile & Space Company  
P. O. Box 1103 West Station  
Huntsville, Alabama

Dr. Robert Baker  
113A 30th Street  
Manhattan Beach, California

Dr. M. L. Anthony  
Space Flight Technology  
Mail No. A-153  
The Martin Company  
Denver 1, Colorado

Mr. Charles F. Pontious  
Guidance & Navigation Program  
Office of Advanced Research & Technology  
Code: REG  
National Aeronautics and Space Administration  
Washington 25, D. C.

Dr. John W. Drebing  
Manager, Systems Analysis  
Advanced Projects Laboratories  
Space Systems Division  
Aerospace Group  
Hughes Aircraft Company  
El Segundo, California

Dr. Joseph F. Shea  
Apollo Program Manager  
Apollo Spacecraft Program Office  
Manned Spacecraft Center  
Houston, Texas

Dr. William A. Mersman, Chief  
Electronic Machine Computing Branch  
Ames Research Center  
Moffett Field, California

Dr. Siegfried J. Gerathewohl  
Bioscience Program Manager  
Manned Space Science Program  
Code SM, NASA  
Washington, D. C.

Dr. Herman M. Dusek  
A. C. Research and Development  
A. C. Spark Plug  
Electronics Division of  
General Motors  
950 N. Sepulveda Blvd.  
El Segundo, California

Dr. D. M. Schrello  
Director, Flight Sciences  
Space and Information Systems Division  
North American Aviation, Inc.  
12214 Lakewood Boulevard  
Downey, California

Mr. Howard S. London  
Bellcomm, Inc.  
1100 17th Street, N.W.  
Washington 6, D. C.

Mr. Howard Haglund  
Jet Propulsion Laboratory  
4800 Oak Grove Drive  
Pasadena 3, California

Mr. Hewitt Phillips  
Langley Research Center  
Hampton, Virginia

Mr. H. A. McCarty  
Space and Information Systems Division  
North American Aviation, Inc.  
12214 Lakewood Boulevard  
Downey, California

Mr. Frank J. Carroll  
Equipment Division  
Systems Requirement Department  
Raytheon Company  
40 Second Avenue  
Waltham, Massachusetts

Chrysler Corporation Missile Division  
Sixteen Mile Road and Van Dyke  
P. O. Box 2628  
Detroit 31, Michigan  
ATTN: Mr. T. L. Campbell or Mr. R. J. Vance (2)  
Dept. 7162  
Applied Mathematics

Dr. I. E. Perlin  
Rich Computer Center  
Georgia Institute of Technology  
Atlanta, Georgia

Dr. D. F. Bender (2)  
Space Sciences Laboratory  
Space and Information Systems Division  
North American Aviation, Inc.  
Downey, California

Dr. Daniel E. Dupree (15)  
Project Leader  
Department of Mathematics  
Northeast Louisiana State College  
Monroe, Louisiana

Dr. Steve Hu (10)  
Northrop Corporation  
Box 1484  
Huntsville, Alabama

Dr. William Nesline (3)  
Raytheon Company  
Sudbury, Massachusetts

Mr. Theodore N. Edelbaum  
Senior Research Engineer  
Research Laboratories  
United Aircraft Corporation  
East Hartford, Connecticut

Dr. M. G. Boyce (3)  
Department of Mathematics  
Vanderbilt University  
Nashville, Tennessee

Mr. Carl B. Cox  
Organization 2-5700  
Mail Stop 22-85  
The Boeing Company  
P. O. Box 3707  
Seattle, Washington 98124

Mr. Dave Engels  
Space and Information Systems Division  
Dept. 595/720  
North American Aviation, Inc.  
Downey, California

Mr. Dale B. Ruhmel  
Staff Associate  
Aeronautical Research Associates of Princeton, Inc.  
50 Washington Road  
Princeton, New Jersey

Dr. Raymond Rishel  
Mathematical Analysis Staff  
Organization 2-5330  
Mail Stop 89-75  
Boeing Company  
P. O. Box 3707  
Seattle, Washington

Dr. Rudolf Hermann, Director  
University of Alabama Research Institute  
4701 University Avenue, N.W.  
Huntsville, Alabama

Dr. S. H. Lehnigk  
Physical Sciences Laboratory  
Army Missile Command  
Redstone Arsenal, Alabama

Mr. R. J. Hayes  
Space Guidance Laboratory  
Electronics Research Center  
Cambridge, Massachusetts 02139

Mr. Harold Chestnut  
1 River Road  
Schenectady, New York



## DISTRIBUTION LIST

### INTERNAL

DIR, Mr. Williams	R-AERO, Mr. Baker (50)
R-FP, Dr. Ruppe	Mr. LeMay
I-I/IB-DIR, Col. James	Mr. Ingram
R-P&VE, Mr. Swanson	Mr. Herring
Mr. T. Miller	Mr. Powers
Dr. Krause	Mr. Causey
	Mr. Lovingood
	Mr. Blair
R-ASTR, Mr. Brandner	R-COMP, Dr. Arenstorf
Mr. Moore	Mr. Davidson
Mr. Richard	Mr. Harton
Mr. Gassaway	Mr. Schollard
Mr. Scofield	Mr. Seely
Mr. Brooks	Mr. Reynolds
Mr. Hosenthien	Mr. Calhoun
Mr. Woods	Dr. Rosen
Mr. Digesu	Dr. Andrus
Mr. R. Hill	AST-S, Dr. Lange
Dr. R. Decher	R-DIR, Dr. McCall
Mrs. Neighbors	Mr. Weidner
R-AERO, Dr. Geissler	MS-T, Mr. Bland (5)
Mr. Jean	MS-IP, Mr. Ziak
Dr. Speer	MS-IPL (8)
Mr. deFries	MS-H
Mr. Cummings	CC-P
Dr. Sperling	DEP-T
Dr. Heybey	I-RM-M
Mr. Thomae	
Mr. Hart	
Mr. Winch	
Mr. Tucker	
Mrs. Chandler	
Mr. Kurtz	
Mr. Stone	
Mr. Teague	
Mr. McNair	
Mr. Dearman	
Mr. Schwaniger	
Mr. Lisle	

## DISTRIBUTION

### EXTERNAL

Dr. W. A. Shaw (10)  
Mechanical Engineering Department  
Auburn University  
Auburn, Alabama

Mr. Samuel Pines (10)  
Analytical Mechanics Associates, Inc.  
941 Front Street  
Uniondale, New York

Mr. J. W. Hanson (20)  
Computation Center  
University of North Carolina  
Chapel Hill, North Carolina

Mr. Richard Hardy  
Mail Stop AG-05  
Boeing Company  
P. O. Box 1680  
Huntsville, Alabama

Mr. Hans K. Hinz (4)  
Research Department  
Grumman Aircraft Engineering Corporation  
Bethpage, Long Island, New York

Dr. Robert Novosad  
Mail Stop #A127  
Martin Company  
P. O. Box 179  
Denver, Colorado

Dr. Dahlard Lukes  
Military Products Group  
Aeronautical Division  
Mail Stop #340  
Minneapolis-Honeywell Regulator Company  
Minneapolis, Minnesota

Mr. Harry Passmore (3)  
Hayes International Corporation  
P. O. Box 2287  
Birmingham, Alabama

Mr. Daniel B. Killeen  
Computer Laboratory  
Norman Mayer Bldg.  
Tulane University  
New Orleans, Louisiana 70118

Dr. Bernard Friedland  
Staff Scientist-Control  
General Precision, Inc.  
Research Center  
Little Falls, New Jersey

Mr. Robert M. Williams  
Chief of Guidance Analysis  
General Dynamics/Astronautics  
Mail Zone 513-0  
P. O. Box 166  
San Diego 12, California

Mr. Ralph W. Haumacher  
A2-863: Space/Guidance & Control  
Douglas Aircraft Corporation  
3000 Ocean Park Blvd.  
Santa Monica, California

Mr. Gary P. Herring  
Chrysler Corporation  
HIC Building  
Huntsville, Alabama

Dr. W. G. Melbourne  
Jet Propulsion Laboratory  
4800 Oak Grove Drive  
Pasadena 3, California

Mr. Myron Schall  
Supervisor, Powered Trajectory  
Space and Information Division  
North American Aviation, Inc.  
12214 Lakewood Blvd.  
Downey, California

Mr. S. E. Cooper  
Space and Information Division  
Dept. 41-697-610, TR-148  
North American Aviation, Inc.  
12214 Lakewood Blvd.  
Downey, California

Dr. George Leitmann  
Associate Professor, Engineering Science  
University of California  
Berkeley, California

Dr. R. P. Agnew  
Department of Mathematics  
Cornell University  
Ithaca, New York

Dr. Jurgen Moser  
Professor of Mathematics  
Graduate School of Arts and Science  
New York University  
New York, New York

Dr. Lu Ting  
Professor of Aeronautics and Astronautics  
New York University  
University Heights,  
Bronx 53, New York

Mr. Clint Pine  
Department of Mathematics  
Northwestern State College  
Natchitoches, Louisiana

Dr. John Gates  
Jet Propulsion Laboratory  
4800 Oak Grove Drive  
Pasadena 3, California

Auburn Research Foundation (2)  
Auburn University  
Auburn, Alabama

Grumman Library  
Grumman Aircraft Engineering Corporation  
Bethpage, Long Island, New York

Jet Propulsion Laboratory Library  
4800 Oak Grove Drive  
Pasadena 3, California

Scientific and Technical Information Facility (25)  
ATTN: NASA Representative (S-AK/RKT)  
P. O. Box 5700  
Bethesda, Maryland

NASA Ames Research Center (3)  
Mountain View, California  
ATTN: Librarian

Dr. Dirk Brouwer  
Yale University Observatory  
Box 2023, Yale Station  
New Haven, Connecticut

Dr. Imre Izsak  
Smithsonian Institution Astrophysical Observatory  
60 Garden Street  
Cambridge 38, Massachusetts

Dr. Peter Musen  
Goddard Space Flight Center  
National Aeronautics and Space Administration  
Greenbelt, Maryland

Mr. Jerome S. Shipman  
Manager, Mission & Mathematical Analysis Group  
Federal Systems Division - IBM  
6702 Gulf Freeway  
Houston 17, Texas

Dr. Henry Hermes  
Assistant Professor  
Division of Applied Math  
Brown University  
Providence 12, Rhode Island

Mr. Curt J. Zoller  
Manager, Guidance and Control  
Information Systems  
North American Aviation, Inc.  
12214 Lakewood Blvd.  
Downey, California

Keith Krusemark  
Senior Engineer Analyst  
Rm. 1167  
Aerospace Industrial Park  
General Electric Company  
Daytona Beach, Florida

Mr. W. J. Orser  
Department of Mathematics  
Lehigh University  
Bethlehem, Pennsylvania

Dr. J. C. Eaves  
Department of Mathematics and Astronomy  
University of Kentucky  
Lexington, Kentucky

Mr. C. H. Tross, President  
Council for Technological and Industrial Development  
P. O. Box 931  
Rancho Santa Fe, California

Mr. Alan L. Friedlander  
Research Engineer  
Guidance and Control Section  
IIT Research Center  
10 W. 35th Street  
Chicago 16, Illinois

Dr. G. J. Etgen, Code RRA  
National Aeronautics and Space Administration  
Washington, D. C. 20546

Mr. Donald Jezewski  
Guidance Analysis Branch  
Spacecraft Technology Division  
Manned Spacecraft Center  
Houston, Texas

Mr. David Thomas  
NASA - Langley Research Center  
Space Mechanics Division  
Langley Station  
Hampton, Virginia

Mr. E. J. Luedtke, Jr.  
Autonetics  
Division of North American Aviation  
3322 Memorial Parkway  
Huntsville, Alabama

Mr. R. K. Squires  
Code 640  
Goddard Space Flight Center  
Greenbelt, Maryland

Dr. V. G. Szebehely  
Yale University Observatory  
Box 2034, Yale Station  
New Haven, Connecticut

Mr. Lembit Pottsepp  
Dept. A-260  
Douglas Aircraft Company  
3000 Ocean Park Blvd.  
Santa Monica, California

Dr. Ronald L. Drake  
Electric Engineering Department  
Drexel Institute of Technology  
Philadelphia, Pennsylvania

Dr. T. J. Pignani  
Department of Mathematics  
East Carolina State College  
Greenville, North Carolina

Dr. Robert W. Hunt  
Department of Mathematics  
Southern Illinois University  
Carbondale, Illinois

Dr. Charles C. Conley  
Department of Mathematics  
University of Wisconsin  
Madison, Wisconsin

Dr. S. Sherman (5)  
Applied Mathematics Subdivision  
Research Division  
Engineering Department  
Republic Aviation Corporation  
Farmingdale, Long Island, New York

Mr. Robert Silber  
Department of Mathematics  
Southern Illinois University  
Carbondale, Illinois

Mr. Robert Glasson  
Bendix Systems Division  
The Bendix Corporation  
3322 Memorial Parkway South  
Huntsville, Alabama

Mr. Wes Morgan  
Mail Stop AF-74  
The Boeing Company  
P. O. Box 1680  
Huntsville, Alabama

Mr. J. S. Farrior  
Lockheed  
P. O. Box 1103 West Station  
Huntsville, Alabama

Douglas Aircraft Corporation  
3000 Ocean Park Blvd.  
Santa Monica, California  
ATTN: R. E. Holmen A2-263  
Guidance & Control Section

Dr. Byron D. Tapley  
Department of Aerospace Engineering  
University of Texas  
Austin, Texas

Mr. Yoshihide Kozai  
Smithsonian Institution Astrophysical Observatory  
60 Garden Street  
Cambridge 38, Massachusetts

Dr. Rudolph Kalman  
Research Institute for Advanced Study  
7212 Bellona Avenue  
Baltimore 12, Maryland

Mr. Ken Kissel  
Aeronautical Systems Division  
Applied Mathematics Research Branch  
Wright-Patterson Air Force Base  
Dayton, Ohio

Mr. Jack Funk  
Manned Spacecraft Center  
Flight Dynamics Branch  
National Aeronautics and Space Administration  
Houston, Texas

Dr. J. B. Rosser  
Department of Mathematics  
Cornell University  
Ithaca, New York

Office of Manned Space Flight  
NASA Headquarters  
Federal Office Building #6  
Washington 25, C. D.  
ATTN: Mr. Eldon Hall

Mr. Bryan F. Dorlin  
Theoretical Guidance & Control Branch  
NASA-Ames Research Center  
Moffett Field, California



Lt. Col. R. A. Newman  
Air Force Space Systems Division  
M-SSVH  
Bldg. 5250  
Redstone Arsenal, Alabama

Dr. Paul Degrarabedian  
Astro Science Laboratory  
Building G  
Space Technology Laboratory, Inc.  
One Space Park  
Redondo Beach, California

Dr. Ray Wilson  
OART - Code RRA  
National Aeronautics and Space Administration  
Washington 25, D. C.

Dr. Joseph W. Siry  
Theory & Analysis Office (547)  
Data Systems Division  
Goddard Space Flight Center  
Greenbelt, Maryland 20771

Mr. Joe Mason  
Space and Information Systems Division  
Department 41-595-720  
North American Aviation, Inc.  
12214 Lakewood Boulevard  
Downey, California

Douglas Aircraft Corporation  
3000 Ocean Park Blvd.  
Santa Monica, California  
ATTN: Mr. Joe Santa  
A2-863

Mr. J. B. Cruz, Jr.  
Research Associate Professor  
Coordinated Science Laboratory  
Urbana, Illinois

Dr. E. B. Lee  
Department of Electrical Engineering  
University of Minnesota  
Minneapolis, Minnesota

NASA Flight Research Center (2)  
Edwards Air Force Base, California  
ATTN: Librarian

NASA Goddard Space Flight Center (2)  
Greenbelt, Maryland  
ATTN: Librarian

NASA Langley Research Center (2)  
Hampton, Virginia  
ATTN: Librarian

NASA Launch Operations Directorate (2)  
Cape Kennedy, Florida  
ATTN: Librarian

NASA Lewis Research Center (2)  
Cleveland, Ohio  
ATTN: Librarian

NASA Manned Spacecraft Center (2)  
Houston 1, Texas  
ATTN: Librarian

NASA Wallops Space Flight Station (2)  
Wallops Island, Virginia  
ATTN: Librarian

Space Flight Library (4)  
University of Kentucky  
Lexington, Kentucky

University of Kentucky Library (10)  
University of Kentucky  
Lexington, Kentucky

Mr. Jules Kanter  
Guidance & Navigation Program  
Office of Advanced Research & Technology  
Code: REG  
National Aeronautics and Space Administration  
Washington 25, D. C.